Calibration and combination of seasonal climate predictions in tropical and extratropical regions

by

Luis Ricardo Lage Rodrigues

Advisors: Prof. Francisco Javier Doblas-Reyes

Dr. Caio Augusto dos Santos Coelho

Tutor: Prof. Ileana Bladé

PhD Program: Physics

Research Line: Astronomy and Meteorology (2010-2015)

October 2015

This thesis is dedicated to my mother.

First, they ignore you,
then they laugh at you,
then they fight you,
then you win.
Mahatma Gandhi

Nós fomos, ao longo do tempo, objeto de olhares extremamente arrogantes.
Dilma Rousseff

Acknowledgements

More than a decade has passed, but I remember as if it had just happened yesterday. I was standing on a line waiting my turn to apply for a written test to join the university[1]. On that day, the line was long as the application deadline was approaching and Brazilians have the bad habit of leaving everything to the last time. I remember to be talking to a colleague about my uncertainties on what bachelor degree to choose. As hard as it may seem, but I was not sure which career to follow just few minutes before applying to the university. I had three options in mind: physics, mathematics and computer sciences. The colleague I referred to previously made an unexpected suggestion "why don´t you apply for meteorology?". He explained his suggestion: "first, less competition (i.e. less people applying for this career) means that it would be easier for you to be accepted by the university, and second, the first two academic years would be similar whether you choose meteorology, physics, mathematics, computer sciences or engineer, for instance. If you join the university you would have more time to think about the best career for you, and then transfer from one department to another". I acknowledged his suggestion and I was accepted to join the Department of Meteorology of the Federal University of Alagoas (UFAL). He also applied for meteorology, but failed the written test. When I think of what I have lived and achieved, I realize that life behaves in a chaotic fashion, just like the atmosphere: a very short conversation many years ago changed completely the course of my whole life.

I found a very interesting environment for learning at the Department of Meteorology of UFAL, despite all limitations a university with limited budget might have. I rapidly started liking studying meteorology as I could learn many things I liked, including physics, mathematics and computer sciences. That is, I no longer needed to choose one of them. I studied very hard, achieving good grades and, in my second academic year, I was awarded with a research scholarship that was renewed each year until my graduation. Just a few hours after my graduation, I moved to Sao Paulo to start a master´s degree at the Brazilian National Institute for Space Research (INPE). Only outstanding students are accepted for graduate studies at INPE and I was one of the lucky ones. The experience at INPE was very enriching to me not only because it has a more robust infrastructure than UFAL, but mainly because the contact I had with very smart students and scientists from Brazil and abroad. All that said I would like start acknowledging my former professors, colleagues and friends from UFAL and INPE. I also appreciate the Brazilian government for having granted me with research scholarships since my second academic year at UFAL until my graduation at INPE. Without their help, I would not have reached until here.

While I was doing a master at INPE, I took a one-month holiday to study English in Canada. There I found students from all over the world and many of them from Latin American Spanish-speaking countries. Interestingly I could understand some of their Spanish, but they could hardly understand anything I said in Portuguese. I remember to

[1] In Brazil, students must take a written test to join a bachelor degree in most universities.

be visiting one touristic attraction, during one of the weekends, when our tour guide gave us some information written in English and Spanish. “Even though the official languages of Canada are English and French; the information provided in this pamphlet is written in English and Spanish because of the importance of the Spanish language in the world…so many people speak it”, he said. As soon as I was back to Brazil, I enrolled myself in a foreign language school and started learning Spanish. This was just a few months before receiving an email from Prof. Francisco Javier Doblas-Reyes informing me that I had been accepted to do my PhD at the Catalan Institute for Climate Sciences (IC3) in Barcelona!

I worked very hard during my PhD and produced good results: two papers published in and one submitted to international journals as first author, one book chapter published as first author and two papers published in international journals as co-author. I also had the opportunity to attend advanced training schools and scientific events in some of the most prestigious climate centers in the world, such as the NCAR (USA), ECMWF (UK), ICTP (Italy), Météo-France (France), APCC (South Korea) and SMHI (Sweden). Finally, I applied last year for and successfully got a permanent position as a public servant at INPE, which I started working since March of the current year. I am confident that my experience in Barcelona helped me get the position. For all that, I have no word that could fully express my gratitude to my PhD adviser, Prof. Francisco (Paco) Javier Doblas-Reyes, for having accepted and advised me during my PhD thesis at IC3. I will be always grateful to him for everything I learned at IC3.

I would like also to express my gratitude to my co-adviser, Dr. Caio Augusto dos Santos Coelho, for having helped me in many instances of my thesis, including teaching me about climate prediction and forecast verification. It is interesting to mention that my co-adviser was co-advised by my adviser while the former was a PhD student at the University of Reading (UK) and the later a researcher at the ECMWF.

I really appreciate my tutor, prof. Ileana Bladé, and the president of the defense committee of my PhD thesis, prof. Bernat Codina Sánchez, for all their support with the administrate issues at the University of Barcelona (UB). I also thank the administrative staff of the Faculty of Physics of the UB for their support.

Perhaps, the most important thing in a PhD thesis is the scientific question one is attempting to address. In this regard, I also appreciate Paco for having provided me with the interesting question about uncertainty quantification in seasonal climate prediction. I learned a lot from him about this and many other topics and considerably improved my scientific skills (information and data organization, written and oral communication, and so on). Besides, I was inserted in his research group, the Climate Forecasting Unit (CFU). The CFU had not only a good computer system, but also a very good staff. Here I take the occasion to appreciate all the support I have received from the excellent CFU IT team:

Oriol Mula, Jordi Peralta (former member), Muhammad Asif and Domingo Manubens. They helped me sort IT issues out whether I was working in the office or remotely from a faraway country. The CFU administrative staff (Mar Rodríguez and Gabriela Tarabanoff) also helped a lot with all the bureaucracy I faced during my time there. Thanks a lot to all of you!

I learned a lot from the talented CFU postdoctoral researchers. Thanks very much Virginie Guémas, Javier Garcia-Serrano and Salvador Pueyo for teaching me many things. I acknowledge my CFU PhD colleagues Aida Pintó and Danila Volpi. Thank you to all the former and current members of the CFU!

I would like to acknowledge the Spanish Ministerio de Economía y Competitividad (MINECO) through IC3 and the European Union's Seventh Framework Programme (FP7) Seasonal-to-decadal climate Prediction for the improvement of European Climate Services (SPECS) project for supporting my PhD. This study was supported by the Spanish MINECO-funded RUCSS project (CGL2010-20657), the European Union's FP7-funded QWeCI (ENV-2009-1-243964) and SPECS (GA 308378) projects, and the Catalan Government. My trips to attend scientific events in distant countries were funded by NCAR, ICTP, APCC, SPECS and the World Meteorological Organization (WMO).

I would like to thank NOAA, NCEP, IRI and NCAR personnel in creating, updating and maintaining the NMME archive. The NMME project and data dissemination is supported by NOAA, NSF, NASA and DOE. I also wish to thank NOAA for providing access to the GHCN, ERSST and GPCP datasets, ECMWF for providing access to the ERA-Interim dataset, DWD for providing access to the GPCC dataset, the UK Met Office for providing access, through the BADC system, to the HadISST dataset and Météo-France and ECMWF for making available their seasonal prediction hindcasts.

Calibration and combination of seasonal climate predictions in tropical and extratropical regions

Abstract

Current technology allows the proliferation of multiple forecast systems developed by different research institutions from all over the world. However, most decision makers need a reliable probabilistic prediction instead of a set of predictions to take an action given the probability of an event to occur. Several studies have shown that the merging of predictions derived from several forecast systems with equal weights yields on average better predictions than the best single forecast system. This approach has been referred to as the simple multimodel (SMM). Nevertheless, none of these studies has shown the existence of a combination method that systematically produces the best predictions. Therefore, this thesis aims at applying different statistical techniques to combine predictions derived from different statistical and dynamical forecast systems to assess whether the performance of the SMM can be improved. These techniques combine the predictions assigning unequal weights to the different forecast systems based on their past performance. A unique feature of this study is the broad nature of the forecast quality assessment, performed using multiple deterministic and probabilistic verification measures and the same verifying observations. This allows comparing the predictions produced by the different combination methods and forecast systems in a coherent way. Besides, most of the forecast systems used in this study are either publicly available or could be easily implemented by the user. This thesis focuses on seasonal prediction of sea surface temperature (SST), near-surface temperature and precipitation in tropical and extratropical regions. It is shown that the predictions of the SMM are often better than the combination methods that assign unequal weights. The difficulty in the robust estimation of the weights due to the small samples available is one of the reasons that limit the potential benefit of the combination methods that assign unequal weights. However, some of the results illustrate under which conditions combination methods that assign unequal weights improve with respect to the SMM predictions. For instance, the combination methods that assign unequal weights improve over the SMM predictions when only a fraction of all single forecast systems have skill as shown for some of the predictions of SST. On the other hand, it is shown that there are cases when combining many forecast systems does not lead to improved forecasts when compared to the best single forecast system. This suggests that a multimodel approach is not necessarily better than a highly skillful forecast system, which highlights the importance of continuously assessing the forecast quality for the specific application of the user.

Key words: Climate prediction, forecast verification, uncertainty quantification

Calibración y combinación de predicciones climáticas estacionales en regiones tropicales y extratropicales

Resumen

La tecnología existente permite la proliferación de varios sistemas de predicción, desarrollados por diferentes instituciones de investigación de todo el mundo. Sin embargo, la mayoría de los tomadores de decisión generalmente necesitan una única predicción probabilística fiable para tomar una acción dada la probabilidad de ocurrencia de un evento. En este sentido, varios estudios han demostrado que la combinación de predicciones derivadas de varios sistemas de predicción resulta, en promedio, en una mejor predicción cuando se compara con la predicción del mejor sistema de predicción. Esto ocurre, entre otros motivos, porque la utilización de varios sistemas de predicción es una manera de cuantificar la incertidumbre inevitable asociada a las aproximaciones utilizadas en la construcción de los sistemas de predicción. No obstante, ninguno de estos estudios ha demostrado la existencia de un método de combinación que produzca las mejores predicciones. Por lo tanto, esta tesis tiene el objetivo de aplicar diferentes técnicas estadísticas para combinar predicciones climáticas estacionales derivadas de diferentes sistemas de predicción. Algunas de estas técnicas ponen pesos desiguales a las predicciones derivadas de los diferentes sistemas de predicción, teniendo en cuenta su calidad en un período pasado. Es decir, las predicciones derivadas de los mejores sistemas de predicción reciben más pesos en la combinación. Una de estas técnicas, conocida como "simple multimodel" (SMM), combina todos los sistemas de predicción sin poner pesos, considerando que tienen la misma calidad.

Sistemas de predicción tanto estadísticos como dinámicos son considerados en este estudio. Los sistemas de predicción estadísticos se basan en el uso de regresión lineal simple y se emplean como benchmark para la comparación con los sistemas de predicción dinámicos más sofisticados. Nueve sistemas de predicción dinámicos son usados en este estudio; entre ellos, dos del proyecto europeo EUROSIP y siete del proyecto norteamericano North America-Multimodel Ensemble (NMME). Los sistemas de predicción dinámicos funcionan en modo operativo o cuasi-operativo y muchos de ellos mantienen sus predicciones disponibles públicamente. Un punto importante de este estudio es el amplio carácter de la verificación de la calidad de las predicciones, ya que se usan varias métricas deterministas y probabilísticas. Además, el método de "bootstrap" no-paramétrico es usado para cuantificar las incertidumbres en los cálculos de las métricas de verificación. Los resultados de esta tesis se dividen en tres partes:

- **Predicción de la temperatura de la superficie del mar (TSM)**

En la primera parte, seis métodos estadísticos son usados para combinar predicciones de índices climáticos de la temperatura de la superficie del mar (TSM) derivadas de tres sistemas de predicción dinámicos y uno estadístico. Las combinaciones estimadas usando un método bayesiano, conocido como "forecast assimilation" (FA; Stephenson et al., 2005), son comparadas con las combinaciones estimadas aplicando los métodos de regresión lineal múltiple descritos en Doblas-Reyes et al. (2005) y el SMM. Se consideran

predicciones de anomalías mensuales de los índices de TSM para todos los meses del año y con una antelación de hasta seis meses, el tiempo máximo para dos de los sistemas de predicción dinámicos usados en esta tesis. La verificación de las predicciones de los índices de TSM en tres océanos tropicales (Pacifico, Atlántico e Indico) se estima aplicando uno índice determinista (coeficiente de correlación) y tres índices probabilistas ("Brier skill score" y dos de sus subcomponentes) para dos eventos de probabilidad (índice de TSM por encima de la mediana y del cuartil superior). De esta manera, se consideran más de 15000 casos diferentes en este capítulo. Como en estudios anteriores, la calidad de las predicciones varía con el mes de inicio, la antelación, región e índice de verificación, de modo que ninguno de los sistemas de predicción es el mejor en todos los casos. También se constató que el sistema de predicción estadístico simple es un buen sistema de control y que algunas veces supera la calidad de los sistemas de predicción dinámicos. Se verificó que las predicciones del SMM son frecuentemente mejores que las que derivan de métodos de combinación que asignan pesos desiguales a los sistemas de predicción. La dificultad a la hora de estimar pesos robustos, debido sobre todo a las pequeñas muestras disponibles, es una de las razones que limita la robustez de las medidas que estiman el beneficio relativo de los métodos de combinación. Sin embargo, hay algunas situaciones en las que los métodos de combinación con coeficientes desiguales son mejores. Esto ocurre, por ejemplo, cuando sólo algunos de los sistemas de predicción tienen calidad predictiva. Los resultados de esta parte de la tesis fueron publicados en la revista Climate Dynamics (Rodrigues et al., 2014a).

- **Predicción de los modos de variabilidad asociados con la precipitación del monzón de África Occidental (MAO)**

En la segunda parte, los métodos de combinación FA y SMM fueron usados para combinar predicciones de los modos de variabilidad asociados con la precipitación del monzón de África Occidental (MAO). Se empleó una nueva metodología para evaluar las variaciones interanuales de la precipitación del MAO, en el que la precipitación mensual se promedia zonalmente entre 10ºW-10ºE antes de estimar los dos modos dominantes de la variabilidad de la precipitación del MAO. El periodo de predicción cubre los meses de Junio a Octubre, un mes antes y un mes después del período de máxima precipitación en la región del MAO (que va de Julio a Septiembre). Los sistemas de predicción dinámicos usados en este estudio permiten calcular las predicciones considerando tres periodos de antelación: cero, uno y dos meses (lo que se consigue con las predicciones iniciadas en Julio, Junio y Mayo respectivamente). Para esta parte se añadieron otros cinco sistemas de predicción dinámicos, que sumados a los tres usados en la primera parte permiten crear un multimodelo de ocho sistemas. Además, otros dos índices de verificación probabilistas se usaron en la evaluación de las combinaciones y de los sistemas de predicción individuales: el "continuous ranked probability skill score" (CRPSS) y el "ignorance skill score". Como en el apartado anterior, pudimos comprobar que en muchos casos en los que los métodos de combinación con coeficientes desiguales son mejores. También se comprobó que el sistema de predicción estadístico produce predicciones de calidad y a veces es difícil que los sistemas de predicción dinámicos lo superen. Se encontró que la combinación de predicciones de varios sistemas de predicción no mejora la predicción del mejor sistema para el caso de los modos principales de variabilidad asociados con la precipitación del MAO. Los resultados de esta parte de la tesis fueron publicados en la revista Journal of Geophysical Research (Rodrigues et al., 2014b).

- **Predicción de la temperatura de la atmosfera próxima a la superficie y la precipitación en Europa y regiones adyacentes**

En la tercera parte, los métodos de combinación FA y SMM se usaron para combinar predicciones de la temperatura atmosférica próxima a la superficie y la precipitación en Europa y regiones adyacentes. Las predicciones se realizan para anomalías mensuales en los meses de verano (entre Mayo y Agosto) e invierno (entre Noviembre y Febrero) inicializadas en los meses de Mayo y Noviembre, respectivamente. En este caso, se consideraron hasta cuatro periodos de antelación. En contraste con los apartados anteriores, en los que los métodos de combinación de predicción se aplican a series temporales (estadística univariada), en este apartado los mismos métodos son usados para combinar predicciones de campos espaciales (estadística multivariada). Se constata que los sistemas de predicción estadísticos basados en regresión lineal no son adecuados para predecir el clima extratropical, a diferencia de lo que ocurre en los trópicos. Los resultados muestran que el método FA es tan bueno o mejor que el SMM cuando varios de los sistemas de predicción simulan bien los principales modos de variabilidad observados. Sin embargo, esto no ocurre frecuentemente en la mayor parte de la región. Los resultados de esta parte fueron enviados recientemente a la edición especial "Climate Variability and Change in the Mediterranean Region" de la revista Global and Planetary Change.

Palabras claves: Predicción del clima, verificación de la predicción, cuantificación de la incertidumbre

Abbreviations

AGCM	Atmospheric General Circulation Model
AMO	Atlantic Multidecadal Oscillation
ARPEGE	Action de Recherche Petite Echelle Grande Echelle
AROC	Area under the Receiver Operating Characteristic
Atl3	SST Atlantic 3 index
BS	Brier Score
BS_c	Brier Score of the Climatology
BS_{GRES}	Generalized resolution term of the Brier Score
BS_{REL}	Reliability term of the Brier Score
BS_{RES}	Resolution term of the Brier Score
BS_{UNC}	Uncertainty term of the Brier Score
BSS	Brier Skill Score
CAM	Community Atmosphere Model
CanAM4	CCCam atmospheric circulation model version 4
CanOM4	CCCma ocean model version 4
CCA	Canonical Correlation Analysis
CCCam	Canadian Center for Climate Modelling and Analysis
CCSM3	Community Climate System Model version 3
CCSM4	Community Climate System Model version 4
CDF	Cumulative Distribution Function
CGCM	Coupled atmosphere-ocean General Circulation Model
CFSv1	NCEP climate forecast system version 1
CFSv2	NCEP climate forecast system version 2
CliPAS	Climate Prediction and its Application to Society
CMC2	Canadian Meteorological Center seasonal forecast system version 2
CMIP3	Coupled Model Intercomparison Project 3
CMIP5	Coupled Model Intercomparison Project 5
CPC	Climate Prediction Center

CRPS	Continuous Ranked Probability Score
$CRPS_{CLIM}$	Continuous Ranked Probability Score of the Climatology
CRPSS	Continuous Ranked Probability Skill Score
DEMETER	Development of a European Multimodel Ensemble System for Seasonal to Inter-Annual Prediction
ECHAM	European Center Hamburg Model
ECMWF	European Centre for Medium-Range Weather Forecasts
ENIAC	Electronic Numerical Integrator and Computer
ENSO	El Niño/Southern Oscillation
EOF	Empirical Orthogonal Function
EOF1	First Empirical Orthogonal Function
EOF2	Second Empirical Orthogonal Function
ERSSTv3b	Extended Reconstructed Sea Surface Temperature analysis version v3b
ESM	Earth System Model
ETI	East Tropical Indian
EUROSIP	European Seasonal to Interannual Prediction
FA	Forecast Assimilation
FAC	FA-Climatology
FAS	FA-Statistical
GCM	General Circulation Model
GEOS5	Goddard Earth Observing System version 5
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecast System
GHCNv2	Global Historical Climatology Network monthly version 2
GloSea4	UK Met Office global seasonal forecast system 4
GPCC	Global Precipitation Climatology Center
GPCP	Global Precipitation Climatology Project
HadAM3	UK Met Office Atmospheric General Circulation Model version 3
HadISSTv1.1	Hadley Center's Global Sea-Ice Coverage and Sea Surface Temperature version 1.1

IFS	Integrated Forecast System
Ign	Ignorance Score
Ign_{CLIM}	Ignorance Score of the Climatology
IgnSS	Ignorance Skill Score
IPCC	Intergovernmental Panel on Climate Change
IRI	International Research Institute for Climate and Society
ITCZ	Intertropical Convergence Zone
JAS	July, August and September
MCA	Maximum Covariance Analysis
MF3	Météo-France seasonal forecast system version 3
MLR	Multiple Linear Regression
MOM3	Modular Ocean Model version 3
MOM4	Modular Ocean Model version 4
NAO	North Atlantic Oscillation
NCEP	National Centers for Environmental Prediction
NEMO	Nucleus for European Modelling of the Ocean
NH	Northern Hemisphere
NMME	North America Multimodel Ensemble
NOAA	National Oceanic and Atmospheric Administration
OGCM	Ocean General Circulation Models
OPA	Océan PArallélisé
PC	Principal Component
PC1	First Principal Component
PC2	Second Principal Component
PCA	Principal Component Analysis
PDF	Probability Distribution Function
POP	Parallel Ocean Program
PROVOST	PRediction Of climate Variations On Seasonal to interannual Time-scales
RPSS	Ranked Probability Skill Score

S3	ECMWF climate forecast system 3
S4	ECMWF climate forecast system 4
SMM	Simple Multimodel
SNA	Subtropical Northern Atlantic
SSE	Sum of the Squared Error
SST	Sea Surface Temperature
USA	United States of America
WAM	West African Monsoon
WMO	World Meteorological Organization
WTI	Western Tropical Indian

Notations

General notations

n	number of training years
N	number of target years
M	number of forecast systems used in the combination
P	number of grid points in the 2.5° grid times the number of forecast systems
Q	number of grid points in the 2.5° grid
^{i}	the superscript i indicates that all years but the ith target year are included in the analysis
$_{i}$	the subscript i refers to the ith target year
$\widehat{\ }^{comb}$	multimodel prediction applying the $\boldsymbol{comb}$ combination technique

Statistical model

$\boldsymbol{x}$	vectors of predictors
$\boldsymbol{X}$	extension of the vector of predictors
$\boldsymbol{y}$	vectors of predictands
$\boldsymbol{a}$	vector of the regression coefficients
a_0	intercept parameter
a_1	slope parameter
$\boldsymbol{\varepsilon}$	vector of residuals
γ^2	predictor variance
$\widehat{\sigma}_0^2$	predicted mean of the squared residuals
$\widehat{\sigma}^2$	predicted variance
$\hat{y}$	predicted mean

Combination methods for univariate predictands

$\widehat{\mathbf{M}}$	matrix of predictors
$\widehat{\mathbf{m}}$	vectors of predictors

$\mathbf{E}$	extension of the matrix of predictors
$\boldsymbol{b}$	vector of the coefficients
b_0	intercept parameter
b_k	kth slope parameter
$\mathbf{S}_{\mathrm{mm}}^2$	covariance matrix of the forecast system´s predictions
$\boldsymbol{F}$	eigenvectors of the covariance matrix of the forecast system´s predictions
$\boldsymbol{\lambda}$	eigenvalues of the covariance matrix of the forecast system´s predictions
$\mathbf{P}$	principal components
$\mathbf{L}$	gain/weight matrix
$\mathbf{G}$	slope parameter of the inverse regression
y_0	intercept parameter of the inverse regression
$\boldsymbol{S}^2$	prediction error covariance matrix of the inverse regression
y_b	predicted mean of the prior distribution
C	predicted standard deviation matrix of the prior distribution

Combination methods for multivariate predictands

$\mathbf{Y}$	spatial-field predictands
$\mathbf{Y}^*$	$\mathbf{Y}$ is reorganized to place the grid points in the rows and the number of training years in the columns
$\hat{\mathbf{F}}$	spatial-field ensemble member of a forecast system for a given target year, variable, target period and lead time. The columns and the rows are, respectively, the longitudes and latitudes for each case
$\hat{\mathbf{S}}$	spatial-field ensemble mean of a forecast system for a given target year, variable, target period and lead time. The columns and the rows are, respectively, the longitudes and latitudes for each case
$\boldsymbol{A}$	$\hat{\mathbf{S}}$ is reorganized to place the grid points and forecast systems in the rows and the number of training years in the columns
$\mathbf{U}$	orthonormal eigenvectors of $\left(\mathbf{Y}^{*\mathrm{T}}\mathbf{A}\right)\left(\mathbf{Y}^{*\mathrm{T}}\mathbf{A}\right)^{\mathrm{T}}$
$\mathbf{V}$	orthonormal eigenvectors of $\left(\mathbf{Y}^{*\mathrm{T}}\mathbf{A}\right)^{\mathrm{T}}\left(\mathbf{Y}^{*\mathrm{T}}\mathbf{A}\right)$
$\mathbf{D}$	diagonal matrix containing the square roots of the eigenvalues of $\mathbf{U}$ or $\mathbf{V}$
$\mathbf{Z}$	left (observations) expansion coefficients

$\mathbf{W}$	right (predictions) expansion coefficients
$\boldsymbol{\mathcal{G}}$	slope parameter of the inverse regression in the MCA space
$\boldsymbol{z_0}$	intercept parameter of the inverse regression in the MCA space
$\boldsymbol{\mathcal{S}}^{2,\mathrm{i}}$	prediction error covariance matrix of the inverse regression in the MCA space
$\boldsymbol{\mathcal{L}}$	gain/weight matrix in the MCA space
$\widehat{\boldsymbol{\Upsilon}}^{FAC}$	FAC predicted mean in the MCA space
$\widehat{\boldsymbol{\mathcal{S}}}^{FAC}$	FAC predicted standard deviation in the MCA space

Forecast quality assessment

r	correlation coefficient
$\hat{p}$	probability forecast
$\hat{\mathrm{p}}_{\mathrm{c}}$	climatological probability forecast
o	observation for a binary event: set to be one if the event happened and zero if it did not happen
L	number of bins used to compute the components of the Brier score
$\mathrm{f}(\chi)$	probability density function of the Gaussian distribution
$\mathrm{F}(\chi)$	cumulative distribution function of the Gaussian distribution
$\mathrm{F}^{*}(\chi)$	generic cumulative distribution function
exp	exponential
ln	natural logarithm
π	mathematical constant

List of Tables

List of Figures

Table of Content

1. Introduction

Several forecast systems have been developed to predict surface climate variables that have a significant impact on human activities, such as near-surface temperature and precipitation. These forecast systems can be classified into two main groups: one that attempts to predict a set of variables based on their relation with another set of variables over a period of time (i.e. statistical forecast systems) and one that applies the laws of physics to predict the evolution of a set of variables (i.e. physically-based numerical forecast systems). A brief overview about these two kinds of forecast systems, including their advantages and limitations are presented in Chapter 2. The definition of the main concepts used in this thesis are also presented in that chapter to make the discussions of the results readable to the general reader.

Currently, the existence of a wide variety of forecast systems in these two groups makes it difficult for users of climate information to choose the best available information to take a decision. Therefore, the application of statistical methods to combine predictions produced by multiple sources of information might be of useful value for users of this kind of information. In addition, it is very important to assess the forecast quality of the different forecast systems to identify the best available information for the user application. The forecast systems used in this thesis as well as the verification techniques used to assess their forecast quality are described in Chapter 3. An important aspect concerning the forecast systems used in this study is that many of them are publicly available on the internet or could be easily estimated by users. Forecast quality assessment is a high dimension problem (e.g. one forecast system could be good to predict precipitation above a median in a certain region, but bad to predict the same variable and region for a different threshold); therefore, several verification measures are used in this thesis in the forecast quality assessment.

Several studies have showed that the combination of several forecast systems yields on average to better forecasts than the best single forecast system. This is because forecast quality varies with variable, region, season and lead time, and therefore, no forecast system proves to be the best one for all situations. However, several questions arise when attempting to combine predictions by several sources of information. For example, what is the best method to combine them? Should all information available be considered in the combination? Is there a way to give more weights to the predictions produced by the better forecast systems based on their past performance in an objective fashion? These are the kind of questions that are considered in this thesis.

1.1. Objectives

The main objective of this thesis is twofold: first, to apply different statistical methods to combine seasonal predictions of climate variables produced by the state-of-the-art statistical and physically-based numerical forecast systems; second, to assess the forecast quality of the resulting combination as well as predictions produced by the individual

forecast systems. This thesis focuses at seasonal climate predictions of sea surface temperature (SST) and precipitation in the tropics and near-surface temperature and precipitation in the extratropics.

This thesis is divided into three main chapters, each considering a different kind of seasonal climate prediction:

1) The first chapter aims at combining three physically-based numerical forecast systems and a simple statistical forecast systems to predict three SST indices over different tropical ocean basins: the Pacific, Atlantic and Indian Oceans. Several combination methods are used to combine the forecast systems and multiple verification statistics are used to assess the forecast quality of the resulting combinations. The idea of working with tropical SST indices (i.e. univariate time series) is to get familiar with the statistical techniques by combining predictions produced by different forecast methods in a simple fashion. The single forecast systems and the combination methods used to achieve this objective are described in Chapter 3 and the results discussed in Chapter 4. The results of this research was published in the journal *Climate Dynamics* (Rodrigues et al., 2014a).

2) The second main chapter applies some of the combination techniques used in Chapter 4 to combine multiple forecast systems to predict the modes of West African monsoon (WAM) rainfall variably. In this chapter, the combination of several forecast systems are performed on the time series associated with the main modes of WAM rainfall variability. Therefore, the combination is dealt with as univariate statistics. To perform this analysis, more forecast systems and verification metrics than the ones used in Chapter 4 are added to the analysis. The forecast systems used for this part of the research as well as the methodology to estimate the WAM modes of rainfall variably are described in Chapter 3. The results of this research are shown in Chapter 5 and published in *Journal of Geophysical Research* (Rodrigues et al., 2014b).

3) Finally, the combination techniques are applied to the monthly near-surface temperature and precipitation predictions over Europe. This final chapter differs from the previous two because the combination of multiple forecast systems are performed on spatial-fields of climate variables. Therefore, this chapter deals with multivariate statistics. An assessment of the forecast quality of multiple statistical and physically-based numerical forecast systems as well as their combinations is discussed in Chapter 6. The results of this research was recently submitted to the special edition *Climate Variability and Change in the Mediterranean Region* of the journal *Global and Planetary Change*.

2. Background

Many human activities are either directly or indirectly affected by weather conditions (Sewell et al., 1968; Trenberth et al., 2000; Koetse and Rietveld, 2009; Doblas-Reyes et al., 2013a). This includes activities as simple as choosing what clothes to wear depending on the near-surface temperature (Sewell et al., 1968) to more complex ones such as the management of aerial, terrestrial or maritime transportation which are highly dependent on wind and precipitation (Koetse and Rietveld, 2009). Other human activities such as agriculture (Trenberth et al., 2000; Coelho and Costa, 2010) and energy production (García-Morales and Dubus, 2007) are also highly dependent on weather conditions. However, these activities are more concerned about *the statistics of weather* over a given period of time rather than the day-to-day weather variations. For instance, information about the mean and the standard deviation of precipitation and near-surface temperature over a period might help farmers decide the best place and/or which plants to grow. Similarly, information about the total amount and annual distribution of solar radiation or wind might help decision makers choose the best place to setup solar or wind power plants.

To make the discussions of the results readable to the general reader, definitions of the main concepts used in this thesis are provided below. Key concepts are highlighted in italic. Both classic and the latest scientific literature will be provided. Besides, each of the main three Chapters of this thesis will review the specific topic literature.

2.1. Earth´s climate system and its variability

According to the World Meteorological Organization (WMO) *climate* is defined as the statistical description of relevant quantities over a period of time. This statistical description could be thought of as the mean and the variability (Wallace and Hobbs, 2006) or more accurately as the probability distribution function (PDF) (Stephenson et al., 2012) of the relevant quantities of the atmosphere and related components of the Earth system. The conventional period of time to define climate as suggested by the WMO is 30 years (Arguez and Vose, 2011). On the other hand, climate prediction usually deals with predictions on any timescales longer than two weeks, the deterministic limit of atmospheric predictability (Lorenz, 1972, 1982). For instance, while it is impossible to predict instantaneous fields of near-surface temperature one month in advance, one could predict its monthly mean one month ahead.

Climate is determined by the energy from the Sun and how it interacts with the atmosphere and the other components of the Earth´s climate system (Wallace and Hobbs, 2006). The amount of energy that reaches the top of the atmosphere is driven by several orbital characteristics such as the *obliquity* (i.e. the tilt of the Earth´s axis of rotation relative to the plane of the Earth´s orbit) and the *eccentricity* (i.e. the shape of the Earth´s orbit around the Sun), which have cycles that last on the order of thousands of years (Wallace and Hobbs, 2006). Currently, the obliquity is 23.5º and the eccentricity is less than 0.02 (i.e. it is almost a circle). These very long cycles allow variations of the solar energy that are mainly periodic and highly predictable (Ghil, 2002). For instance, because of the obliquity and the rotation of the Earth around the Sun it is easy to predict that, on average, winters will be colder than summers.

Only half of the *in*coming *so*lar radia*tion* (i.e. the *insolation*) reaches the Earth´s surface, the other half is either reflected back to space or absorbed by clouds, particles and gases

contained in the atmosphere (Hartmann, 1994). The insolation that reaches the surface has a latitudinal gradient, especially due to the obliquity. This happens because the insolation that falls perpendicular or nearly so close to the equator warms the near-surface temperature more intensely than in higher latitudes of both hemispheres (Hadley, 1735). As a result, a thermal gradient between the equator and the poles is established, which in turn, creates a near-surface pressure gradient with lower pressure near the equator than at higher latitudes. Consequently, a near-surface meridional circulation where wind blows from the higher latitudes to the equator is established. The equatorward near-surface wind coming from both hemispheres is deflected by the Earth´s rotation to become the northeasterly and southeasterly *trade winds* whereas the belt in which trade winds converge is known as the *intertropical convergence zone* (ITCZ) (Wallace and Hobbs, 2006).

Strong upward motion and convective activity characterizes the ITCZ region. In this region, the air rises and reaches the top of the troposphere and then diverges towards the poles of both hemispheres. As the air moves towards higher latitudes transporting both heat and angular momentum it is deflected eastwards due to the Earth´s rotation to become the mid-latitude westerly winds. When moving away from the equator the air gets cooler and denser, sinking at the subtropical region known as *subtropical high belt*. Finally, the air returns to the equator in the form of trade winds closing this thermally direct meridional circulation, known as the *Hadley cell* in recognition of the famous paper written by lawyer and amateur meteorologist George Hadley almost three centuries ago (Lorenz, 1967, 1983; Dima and Wallace, 2003; Schneider, 2006; Wallace and Hobbs, 2006).

About a century after the publication of Hadley´s famous paper, more observations became available and showed that his theories could not fully explain the general circulation of the atmosphere (Lorenz, 1967, 1983; Schneider, 2006). For instance, the observed near-surface wind in the extratropics turned to be in fact westerly and not easterly, as expected from the air traveling from the poles to the equator. Consequently, in this region, near-surface wind actually travels poleward and not equatorward (Ferrel, 1859). One of the points missed by Hadley was the large-scale eddies that are not seen in an idealized symmetric circulation but are responsible for a great amount of atmospheric meridional transport of heat and angular momentum (Lorenz, 1967, 1983; Hartmann, 1994; Schneider, 2006). Hadley also missed a third meridional thermally direct circulation in the polar region, known as the *polar cell*. Thus, the *three circulation cells* in each hemisphere give a simplified but not complete picture of the general circulation of the atmosphere (Schneider, 2006). The climate of the WAM in the tropics and Europe in the extratropics, whose predictability are assessed and discussed in Chapters 5 and 6, respectively, are significantly affected by this meridional circulation.

The meridional circulation of the atmosphere is usually described in terms of zonally-averaged climate variables. However, the global atmospheric circulation is asymmetric among other things because of the distribution of land and ocean. Therefore, a zonal asymmetric circulation must be considered to represent a wider picture of the general circulation of the atmosphere. The zonal circulation was first introduced by Bjerknes (1969) when studying the atmospheric circulation over the equatorial Pacific Ocean (Hastenrath, 1991; Lau and Yang, 2002). This zonal circulation was named the *Walker circulation* because it was considered an important part of the mechanism of the so called *Southern Oscillation* (Bjerknes, 1969), a coherent pattern of pressure, near-surface

temperature and rainfall fluctuations first described by Sir Gilbert Walker (Mock, 1981; Rasmusson and Wallace, 1983). Today, the term Walker circulation is sometimes referred to as the totality of the global tropical zonal circulation (Hastenrath, 1991; Lau and Yang, 2002; Wang, 2005).

In an attempt to develop a seasonal forecast scheme for the Indian monsoon, Sir Gilbert Walker studied contemporaneous and lagged correlation of pressure, near-surface temperature and rainfall in different regions around the globe and different seasons of the year (Mock, 1981). He showed that the global atmosphere has several reproducible patterns of low-frequency variability, among them, one in which the surface pressure in Tahiti and Darwin were anti-correlated, a pattern he named the Southern Oscillation (Mock, 1981; Rasmusson and Wallace, 1983; Roulston and Smith, 2002). Simple statistical forecast schemes are still used in climate prediction as the ones used in this thesis.

As mentioned previously, the near-surface wind in the global tropics is easterly which help explain the SST pattern over the Pacific Ocean (Bjerknes, 1966, 1969). Under normal conditions, near-surface wind takes warmer waters to the western bound of the tropical Pacific Ocean, establishing a region with anomalous warm SST, which is known in the literature as the *warm pool* (Lau and Yang, 2002). This establishes a strong westward (eastward) gradient of SST (pressure) along the tropical Pacific. As a consequence, the Pacific Walker circulation is characterized by near-surface convergence, rising motion and deep convection over western Pacific and subsidence and absence of rainfall near the dateline (Bjerknes, 1966, 1969; Rasmusson and Wallace, 1983; Hastenrath, 1991). Therefore, the Walker circulation is thermally direct.

Bjerknes (1969) showed that an increase in the east-west pressure gradient in the equatorial Pacific is associated with a stronger near-surface westward wind, which would increase the east-west SST gradient. This positive ocean-atmosphere feedback would intensify the Walker circulation over the Pacific Ocean. However, Bjerknes also noted that the opposite could happen as follows: a weaker near-surface westward wind would increase the SST in the eastern bound of the Pacific Ocean, thus decreasing the east-west SST gradient, and as a consequence, weakening the Walker circulation. The warming of the eastern bound of the tropical Pacific SST is one of the fingerprints of a very important climate phenomenon, known as the *El Niño* (Hastenrath, 1991; Trenberth, 1997; Trenberth et al., 2000; Wallace and Hobbs, 2006). The El Niño events, and their counterpart La Niña, are usually referred to in the climate community as the oceanic component of the so called coupled ocean-atmosphere phenomenon known as the *El Niño/Southern Oscillation* (ENSO). The atmospheric component of ENSO, the Southern Oscillation, is defined as the pressure difference between Tahiti and Darwin.

An interesting feature of ENSO is that its influence is not only limited to the climate of the tropical Pacific Ocean where it takes place, but also on distant regions around the globe. Several studies have shown ENSO impacts on near-surface temperature and precipitation over different regions across the globe (Ropelewski and Halpert, 1987; Halpert and Ropelewski, 1992; Rodó et al., 1997; Grimm et al., 1998; Bronnimann, 2007; Chase et al., 2007; Zanchettin et al., 2008; Joly and Voldoire, 2009). Figure 2.1 illustrates the impacts of El Niño, the warm episode of ENSO, on different regions across the globe in boreal winter. Examples of El Niño impacts are wetter and warmer conditions in eastern and central equatorial Pacific Ocean, drier conditions in western equatorial Pacific

Ocean, wetter and cooler conditions in Southern North America, drier and warmer conditions in Northern South America. These are examples of the so-called ENSO *teleconnection pattern*, which is defined as the recurring and persistent large-scale patterns of climate anomalies that extend over vast geographical areas and have a remote origin (Wallace and Gutzler, 1981; Panagiotopoulos et al., 2002; Chase et al., 2007). Of course, every ENSO event is different and so are their impacts, which make their predictions particularly challenging.

A great amount of latent heat that drives the global atmospheric circulation is released in the ascending branch of the Walker circulation over the warm pool (Lau and Yang, 2002). Other sources of latent heat in the atmosphere are located in the Amazon and in tropical Africa. These three sources of heat are considered the three major global sources of heat (Wang, 2005). Changes in these sources of heat could lead to anomaly circulation patterns both locally and in distant regions because the Earth´s climate, as a closed system (i.e. the exchange of matter between Earth and space is almost negligible), attempts to equalize the spatial distribution of heat.

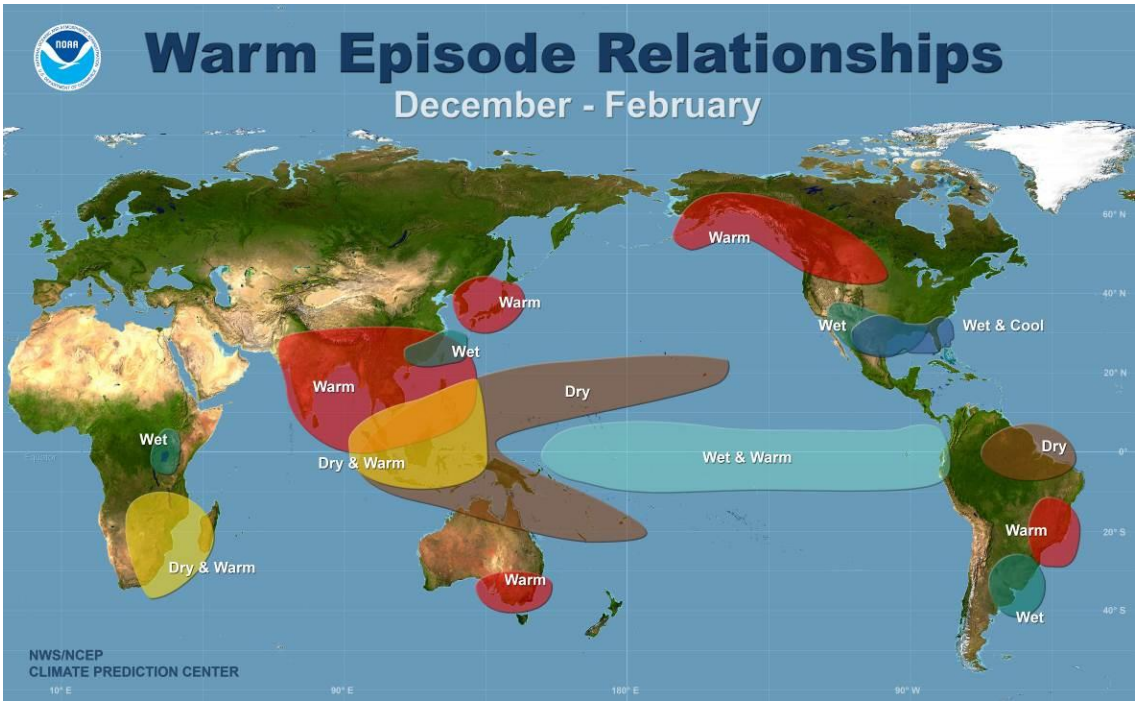

Figure 2.1: Illustration of global impacts of El Niño during the boreal winter. Source: CPC/NCEP/NOAA (http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/ENSO-Global-Impacts/High-Resolution/).

The general circulation of the atmosphere shows significant variability in time as the result of different processes that occur on many spatial and timescales (Wallace and Hobbs, 2006). For instance, a monsoon system, which is the result of an asymmetric heating between the land and the ocean, is characterized by having a well-defined rainy season that takes place during summertime (Hastenrath, 1991). In other words, the seasonal variability in a monsoon system is well established with a wet season in summer and a dry season in winter. However, variability at other timescales also takes place in a monsoon system, such as the ones that occur within a season when active and break

periods occur (i.e. intraseasonal variability) or from year to year when the intensity of the wet season in a given year is stronger or weaker than normal (i.e. interannual variability). In this thesis, we are concerned by the prediction of climate variations on interannual timescales.

ENSO is the most important global teleconnection pattern at interannual timescales (Doblas-Reyes et al., 2013a; Hoskins, 2013). However, for specific regions, other teleconnection patterns are of great importance (Wallace and Gutzler, 1981). For instance, the North Atlantic Oscillation (NAO), a large-scale seesaw phenomenon between the North Atlantic subtropical high and the polar low, is the main source of interannual variability in the circulation over the North Atlantic region (Wallace and Gutzler, 1981; Hurrel, 1995; Wanner et al., 2001; Chase et al., 2007). It influences the climate over North America, Europe and Eurasia (Wanner et al., 2001; Chase et al., 2007). The positive phase of NAO happens when both the subtropical high and the polar low are stronger than average, which increases the meridional gradient of pressure and, as a consequence, creates stronger high-altitude winds and large-scale eddies crossing the North Atlantic Ocean (Figure 2.2). It is usually associated with warmer and wetter winters over Northern Europe and drier winters over Southern Europe and the Mediterranean region (Chase et al., 2007). The opposite is observed during the negative phase of the NAO.

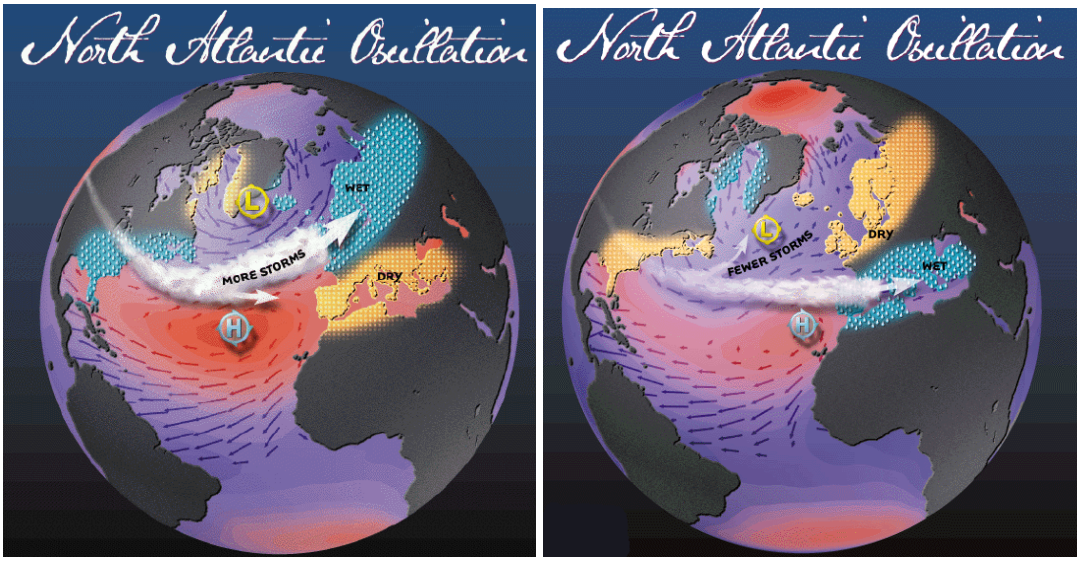

Figure 2.2: Illustration of the positive (left) and negative (right) phases of the NAO. The positive phase of NAO happens when both the North Atlantic subtropical high and the polar low are stronger than average, which increases the meridional gradient of pressure and, as a consequence, creates stronger wind and large-scale eddies crossing the North Atlantic Ocean. This is associated with warmer and wetter winters over Northern Europe and drier winters over Southern Europe. The opposite is observed during the negative phase of the NAO. Source: http://www.ldeo.columbia.edu/res/pi/NAO/.

Understanding the nature of teleconnections and changes in their behavior is a very important step to understand climate variability at the regional level (Trenberth et al., 2007). Therefore, knowledge of the variations of the teleconnection patterns provides a way to increase climate predictability at the regional level as illustrated in the examples of ENSO and NAO. Many teleconnection patterns have been identified, but only a small number of them explain most of the seasonal to interannual climate variability in the

circulation and surface climate variables (Wallace and Gutzler, 1981; Wanner et al., 2001; Trenberth et al., 2007). Characterizing the teleconnection patterns (i.e. using standard teleconnection indices) was attempted by Walker and Bliss (1932) and improved by Wallace and Gutzler (1981) who used more modern and objective approaches to describe them. Since then, several studies have attempted to improve the understanding of the physical mechanisms modulating the teleconnection patterns, and how they affect the climate of different regions (Panagiotopoulos et al., 2002). Today, teleconnection indices are routinely used by different operational climate centers across the globe as important monitoring and forecasting tools (Panagiotopoulos et al., 2002; Chase et al., 2007; Cohen and Jones, 2011; Korecha and Sorteberg, 2013).

2.2. Seasonal climate prediction

According to the WMO, a climate prediction is a probabilistic statement about future climate conditions on timescales ranging from months to decades. Climate prediction does not attempt to forecast which day a given weather situation will take place over a specific region; therefore, it is usually described in terms of statistical properties computed over a period of time (Lorenz, 1982). Seasonal climate prediction falls in a time window between one month and a couple of years and a probabilistic statement is usually formulated in terms of monthly or three-month averages or total values (Doblas-Reyes et al., 2013a).

Climate predictability relies on factors that have a continuous influence over a period of time (Charney and Shukla, 1981; Godard et al., 2001; Doblas-Reyes et al., 2013a; Hoskins, 2013). These factors are usually observed in the components of the climate system that move slowly, such as ocean, land surface and sea ice. The predictability of the atmosphere comes from the interaction with these slow varying components of the climate system. ENSO, in its both phases, is the most important mode of climate variability at seasonal timescales characterized by slowly varying SST in the equatorial Pacific (Goddard et al., 2001; Doblas-Reyes et al., 2013a; Hoskins, 2013).

Some of the approaches currently used to perform climate prediction at the seasonal timescale are described below. The first approach uses statistical relationships estimated from historical observations to infer about future climate conditions, while the second one uses the laws of physics that govern the climate system, quantified in the form of numerical equations. Given the availability of a large number of forecasting methods and the need of users to have a single information to take any action, a third approach arises to combine all available information to estimate a single source of information of the future climate state (Doblas-Reyes et al., 2013a).

2.2.1. Statistical approach

In this approach, statistical methods are used to identify relationships between two sets of variables through the analysis of historical observations (Wilks, 2006; Mason and Baddour, 2008). The first set of variables are the ones to be predicted, often referred to as *predictands*, while the second one are the ones to make the predictions, often referred to as the *predictors*. The simplest forecasting method is to assume that the climate conditions at the time of the forecast will not change, that is, that the predictand will be equal to the predictor. This method is referred to in the literature as the *persistence forecast*. Another simple forecasting method consists in assuming that the predictand will be equal to its

probability distribution over a period of many years. This method is referred to as the *climatological forecast* or simply *climatology*. Persistence and climatological forecasts are often used as a *reference forecast* for evaluating the performance of more sophisticated prediction methods. When a given forecasting method performs better than the reference forecast it is usually said that this forecasting method has *skill* (Wilks, 2006). The climatology is the reference forecast used in this study.

Regression analysis is the one of the most common statistical techniques applied to seasonal forecasting (Mason and Baddour, 2008). Linear regression analysis allows to predict the behavior of a predictand based on the variations of just one predictor (i.e. simple linear regression) or more than one predictor (i.e. multiple linear regression). Regression analysis can also be used when the relationship between predictand and predictor is nonlinear (i.e. nonlinear regression) or when predictand and/or predictor have multiple dimensions (i.e. multivariate regression).

Several assumptions must be met before attempting to use any statistical method in climate forecasting (Mason and Baddour, 2008). The first assumption is that the future climate will behave like it did in the past over a fixed period of time. This might be an issue considering the Earth´s climate is non-stationary (Arguez and Vose, 2011; Doblas-Reyes et al., 2013a). The second assumption is that the historical observations are of high quality (i.e. they are accurate and well distributed spatially and vertically). In reality, this is not always the case despite the fact of recent improvements in the observing system with the use of satellites and automatic weather stations. The lack of necessary observations is still an issue in some regions around the world, particularly in the higher atmosphere and deeper ocean (Le Treut et al., 2007; Hartmann et al., 2013). The third assumption is that the sample size is large enough to represent the true population. This might be an issue in seasonal forecasting given the small sample size available (Goddard et al., 2001, 2013; Coelho et al., 2004; Palmer et al., 2004; Doblas-Reyes et al., 2005, 2009, 2013; Mason and Stephenson, 2008). For seasonal climate forecasts, it is recommended at least 30 years of observations to construct a statistical model (Mason and Baddour, 2008).

Other assumptions must be made when using linear regression models (Wilks, 2006; Mason and Baddour, 2008) as follows:

- the relationship between the predictor(s) and predictand is linear;
- the predictand is normally distributed;
- the residuals from the regression analysis are independently of each other and normally distributed with zero mean and constant variance;
- when there are more than one predictor, two or more of them are not linearly well correlated.

Violation to one of these assumptions might lead to regression coefficients and confidence intervals that are not robust, resulting in bad forecasts.

In spite of the difficulties found to satisfy all these assumptions, there are several advantages of using statistical methods in climate forecasting (Doblas-Reyes et al., 2013a). First, statistical models are based on real-world observations, therefore, they are unbiased. Second, they are usually very simple, which make them both easy to understand and cheap to run (i.e. they do not require expensive computer resources). Third, they may be based purely on statistical relationship without any knowledge about the physical processes, although the choice of predictors usually takes into account some physical

knowledge about the existence of some relationship between predictors and predictands. For instance, one of the first statistical model was derived without a satisfactory physical basis (Walker and Bliss, 1932). In this thesis, simple linear regression is used to predict SST anomalies over three different ocean basins (i.e. the Pacific, Atlantic and Indian Oceans), two modes of WAM rainfall variability, and near-surface temperature and precipitation over Europe.

2.2.2. Dynamical approach

Concurrently with the first attempts to forecast the future atmospheric state using statistical methods, a growing number of scientists started trying to apply the principles of theoretical physics to simulate the behavior of the atmosphere (Roulston and Smith, 2002). Vilhelm Bjerknes was one of the first to suggest the use of physical laws on the weather forecasting problem (Roulston and Smith, 2002; Lynch, 2008). He proposed two necessary steps to successfully perform a physically-based weather forecasting (Bjerknes, 1904):

> *1. A sufficiently accurate knowledge of the initial state of the atmosphere*
>
> *2. A sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another*

The first step (i.e. diagnostic) demands a proper observation network to represent the three-dimensional structure of the atmosphere at a particular time, while the second one (i.e. prognostic) requires a system of nonlinear equations with seven dependent variables and seven independent equations that describe the behavior of the atmosphere (Bjerknes, 1904; Gronos, 2005; Lynch, 2008).

A few years later, mathematician Levis Richardson used his own finite difference method to attempt to solve analytically by hand the system of nonlinear equations proposed by Vilhelm Bjerknes (Lynch, 2008). Richardson´s effort to perform a six-hour weather forecast over two points in central Europe might have taken as much as two years to be accomplished and failed completely (Lynch, 1993, 2008). Posterior analysis suggested that his failure was caused because no smoothing technique was applied to deal with high-frequency atmospheric oscillations, such as the gravity waves, and violation of the numerical stability requirement. A description of the numerical methods used in his attempt was published in the book *Weather Prediction by Numerical Process* (Richardson, 1922). Richardson´s ideas were not taken seriously at his time, but form nowadays the basis of modern numerical weather and climate prediction (Lynch, 2008).

The first successful physically-based numerical weather prediction was only achieved in the 1950s by a group of scientists, led by meteorologist Jule Charney, over the continental United States (Roulston and Smith, 2002; Holton, 2004; Lynch, 2008). The successful prediction was only possible due to a combination of factors (Holton, 2004; Lynch, 2008). First, there was a substantial increase in the network of surface and upper-air observations, which significantly improved the diagnostic step of the prediction (i.e. the initial conditions). Second, more observations allowed a better understanding the physical processes of the atmosphere. At this point, Charney contributed to the advancement of the understanding of the large-scale atmospheric circulation and derived a set of

simplified equations by introducing the geostrophic and hydrostatic approximations that filter out the atmospheric high-frequency oscillations. Third, a better understanding of the numerical stability processes allowed an improved performance of the finite difference method. Finally, the development of the first digital computer, the Electronic Numerical Integrator and Computer (ENIAC), made huge amount of numerical operations in a short period possible. This 24-hour forecast took 24 hours to be computed using ENIAC.

It did not take long before scientists start attempting to apply numerical methods to simulate the behavior of the atmosphere at hemispheric or global scale, giving the foundation of the so called general circulation models (GCMs; Edwards, 2011). Norman Philips was one of the first to employ an atmospheric GCM (AGCM) to model numerically the general circulation of the atmosphere (Holton, 2004). Philips (1956) showed that several aspects of the general circulation of the atmosphere, such as the distribution of surface zonal wind and the poleward transport of energy, could be simulated using an AGCM. Besides Philips´ successful experiments using simplified equations, it was considered necessary to solve the full set of nonlinear equations proposed by Vilhelm Bjerknes half a century earlier to realistically simulate the atmospheric general circulation in AGCMs (Roulston and Smith, 2002; Holton, 2004; Lynch, 2008; Edwards, 2011).

In the following decades, several research groups started more or less independently to build multi-leveled three-dimensional AGCMs based on the set of nonlinear equations (Edwards, 2000). Despite their differences, all modern AGCMs have a core with a system of nonlinear equations, often called primitive equations. These primitive equations solve variables such as temperature, pressure, moisture and the three components of the wind. Many physical processes that are very important for the large-scale atmospheric circulation cannot be explicitly solved because either they occur at very small scale or they are too complex to be solved explicitly numerically or they are not entirely understood. These unresolved physical processes are estimated as a function of the resolved variables in a simplified fashion using a technique known as the *parameterization* (Holton, 2004). Physical processes associated with cloud, radiation and rain are examples of physical processes that are parameterized in numerical models. Parameterizing physical processes is one of the aspects that introduces large amounts of uncertainty in physically-based numerical models (Edwards, 2011 and references therein).

GCMs are used in both weather and climate prediction, but they address the problem in different ways. While weather prediction is often considered an *initial value* problem due to its high sensitivity to the initial conditions of the atmosphere, climate prediction is frequently denoted as a *boundary value* problem because its predictability derives mainly from the slowly varying boundary conditions of the climate system (Goddard et al., 2001; Holton, 2004). Figure 2.3 illustrates how seasonal atmospheric anomalies in the tropical Pacific are insensitive to the initial conditions of the atmosphere (Shukla, 1998). Instead, the underlying boundary conditions play a major role. In this example, an AGCM is used to simulate the December-February seasonal mean rainfall anomalies over the Pacific Ocean using the same underlying observed SST, but two very distinct initial conditions of the atmosphere. It is shown that the two simulations are very similar to each other and to the observed rainfall patterns, except for some regions where deficiencies in the simulations were already known (Shukla, 1998).

As mentioned previously, there are nonlinear feedbacks between the atmosphere and the other components of the climate system, which are usually hard to understand and predict (Ghil, 2002). Perhaps, the most famous example of such nonlinear feedback is the ENSO. Attempting to predict the Earth´s climate without taking into account the interactions between the atmosphere and the difference components of the climate system would eventually lead to bad forecasts (Mason, 2008a). Manabe and Bryan (1969) is often mentioned as one of the first studies to identify the need to develop a forecasting system in which an atmospheric model is coupled to an ocean model. Ocean models have evolved to become ocean GCMs (OGCMs), with a full system of nonlinear equations and parameterizations (McWilliams, 1996). Several forecast centers around the world couple an AGCM and an OGCM to form a fully coupled atmosphere-ocean GCMs (CGCM) that are used as an important source of information for operational climate prediction (Doblas-Reyes et al., 2013a).

Monthly or seasonal mean climate anomalies can be realistically simulated by AGCMs when observed boundary conditions are known (Bengtsson et al., 1993; Shukla, 1998; Goddard and Mason, 2002; Tippet and Giannini, 2006). However, observed boundary conditions are not available when an AGCM is used to forecast the future; therefore, they must be predicted. There are two common approaches used to perform climate prediction using GCMs (Goddard et al., 2001; Doblas-Reyes et al., 2013a). In the first one, the boundary conditions are predicted first by a statistical model, a physically-based numerical model or considering the observed SST anomaly of the month prior to initialization of forecasts is persisted during the forecast period and only then an AGCM is used to perform the forecast of the atmospheric variables. This approach is known as the *two-tier forecast* and is currently rarely employed in operational contexts. In the second approach, the atmosphere and two or more slow varying components of the climate system are explicitly modeled and simulated synchronously and interactively. Thus, the atmosphere not only is influenced by these slow varying boundary conditions, but it also interacts with them as they evolve with time. This approach is known as the *one-tier forecast*.

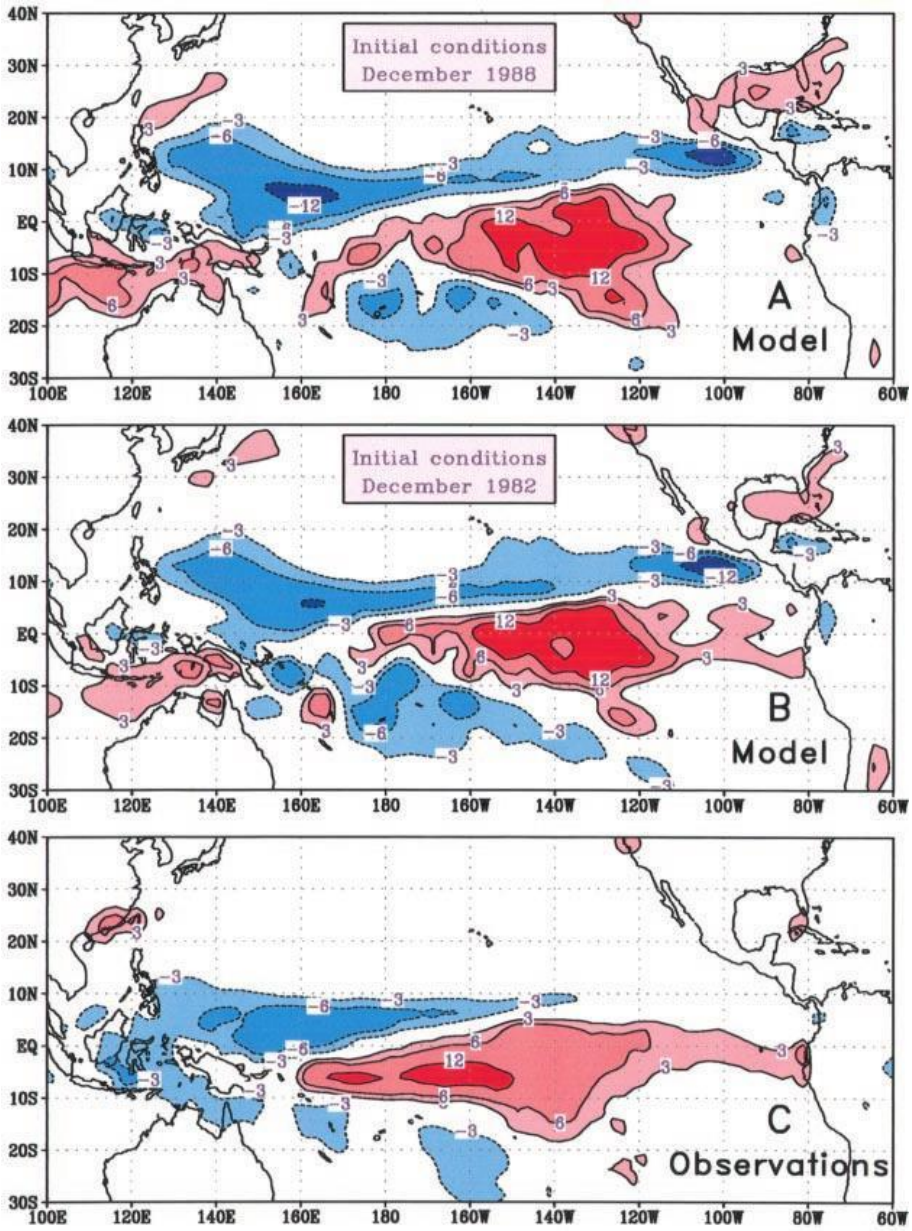

Figure 2.3: Average rainfall anomaly (mm day^{-1}) for January, February, March for two sets of five-model integrations with observed SST in 1982–1983 starting from atmospheric initial conditions in mid-December 1988 (A) and 1982 (B) and observed (C). Source: Shukla (1998).

More recently, CGCMs have been coupled to the other components of the climate system, such as the land surface and the cryosphere, to create what is known as Earth system models (ESM; Edwards, 2011). The latest Intergovernmental Panel on Climate Change (IPCC) reports have used both CGCMs and ESM to simulate *climate change projections* of the climate system (Collins et al., 2013). As mentioned previously, the initial state of the atmosphere does not play the only role in climate prediction; consequently, it was frequently considered to be a boundary layer problem. On the other hand, because current CGCMs and ESMs require that slowly varying components of the climate system must be initialized (Hurrell, 2008) and external forcing might play a role even at seasonal timescale (Doblas-Reyes et al., 2006), climate prediction should be viewed as both an initial value and boundary condition problems (Pielke, 1998). Short-term climate

predictions, where the system can take advantage of the long memory of the initial conditions of the slowly varying components of the climate system, are known as the *climate prediction of the first kind* (Lorenz, 1975; Kirtman et al., 2013). Longer-term climate predictions that are more sensitive to external forcing boundary conditions, such as the variations in the greenhouse gases, are known as the *climate prediction of the second kind* (Lorenz, 1975), today often referred to as climate projection (Collins et al., 2013). Although the limit between climate prediction of the first and the second kind is still under debate, the memory of the initial conditions of the non-atmospheric components of the climate system might play an important role on predictions up to a couple of years in advance (Doblas-Reyes et al., 2013b). For the sake of simplicity, CGCMs and ESMs will be referred to as dynamical forecast systems hereafter.

When compared to seasonal climate prediction performed by statistical models, dynamical forecast systems have two main advantages: first, they are able to predict climate events that have never been observed in detail, and second, they are able to take into account the nonlinear interactions among the different components of the Earth´s climate system. On the other hand, they also have several disadvantages when compared to statistical models. For instance, the system of nonlinear equations used to build dynamical forecast systems need to be simplified to be solved numerically and many important physical processes are parameterized. These two steps introduce uncertainty in the forecasts. Besides, dynamical forecast systems require computer power of orders of magnitude many times higher than that needed to perform seasonal climate prediction with statistical models.

2.2.3. Uncertainty Quantification

The first attempts to perform weather forecasting using dynamical forecast systems date back to the beginning of the 20th century. However, by the 1950s, when the first studies demonstrating the effectiveness of such approach, most meteorologists were still using linear statistical methods to predict the weather (Motter and Campbell, 2013). At that time, meteorologist Edward Lorenz were among those researchers seeking to prove that non-periodic atmospheric processes could not be predicted deterministically a few days in advance using linear statistical methods (Lorenz, 1960). Physically-derived dynamical or nonlinear statistical methods should be used instead.

Edward Lorenz spent a lot of time trying to find a system of nonlinear equations whose solutions would give the simplest example of a deterministic non-periodic flow (Motter and Campbell, 2013). After finding a simplified convective model that met his requirements, Lorenz (1963) showed that, in a system with bounded solutions, slightly different initial states could develop into considerably different states. The sensitive dependence on initial conditions is a behavior of nonlinear systems known as the *deterministic chaos* or simply *chaos*. This property has greatly influenced a wide range of basic sciences, being considered as one of the most important findings of modern science (Motter and Campbell, 2013). Chaos implies that there is always a finite limit to the predictability of a chaotic system. Lorenz (1982) envisioned the limit of two weeks for deterministic forecasts of instantaneous weather patterns, while predictability beyond that limit would be restricted to some properties such as weekly, monthly, and longer-period averages and other statistics. A recent study using a modern dynamical forecast system indicates that this limit envisioned by Lorenz is close to be achieved (Froude et al., 2013).

The dimensional space occupied by the trajectories of the solutions of the simplified convective model mentioned above is known as the *Lorenz attractor* (Figure 2.4) (Slingo and Palmer, 2011). Because of its shape, the Lorenz attractor is frequently referred to as the *butterfly effect* (Motter and Campbell, 2013). Because of chaos, uncertainty must always be accounted for in any weather and climate prediction (Slingo and Palmer, 2011). One of the methods used to take into account uncertainty in the initial conditions is *ensemble forecasting*, where a dynamical forecast system performs several predictions with slightly different initial conditions (Palmer, 2000; Gneiting and Raftery, 2005, Wilks, 2006).

Figure 2.4 shows three ensemble forecast experiments in the Lorenz attractor to illustrate that the predictability of a chaotic system is flow dependent, that is, certain initial states are more predictable than others (Slingo and Palmer, 2011). In the first experiment of the left panel, there is very little divergence or spread among the ensemble members throughout the ten time-steps. All ensemble members point to a change of regime from the left-side regime to the right-side regime. This regime could be thought of as, for example, a dry regime on the left-side of the attractor and a wet regime on the right. Hence, there is large predictability in these initial conditions. In the experiment of the middle panel, the ensemble members start diverging after the time-step four. In this case, even though there is less predictability than in the first case, it is still possible to infer in a probabilistic fashion that it is more likely that there would not be a change of regime as a larger number of ensemble members keep on the left-side of the attractor. In the third experiment of the right panel, the ensemble members diverge completely apart after a few time-steps, resulting in a very small predictability.

Figure 2.4: Ensemble forecasting illustrated by the prototypical Lorenz (1963) model of chaos showing that, in a nonlinear system, predictability is flow dependent. (a) A forecast with high predictability, (b) forecast with moderate predictability and (c) forecast with low predictability. Source: Adapted from Palmer et al. (2005).

Ideally, the number of ensemble members should be large enough to estimate the PDF that represents the true uncertainty in the initial condition (Wilks, 2006). However, limited computing power impose restrictions to the number of ensemble members. For instance, dynamical forecast systems used in seasonal forecasting usually have from ten to fifty ensemble members as the ones used in this thesis, described in Chapter 3. The first step to quantify the uncertainty associated with initial conditions is to incorporate available observations into the dynamical forecast system space, in a process known as *data assimilation* (Wilks, 2006). The combination of predictions and observations is

known as the *analysis*. The set of initial conditions used to sample uncertainty in ensemble forecasting are estimated from perturbations of the analysis of atmospheric variables in weather prediction and ocean, land or ice variables in climate prediction.

There are several methods of perturbation aimed to quantify the initial condition uncertainty in an ensemble forecasting (Wilks, 2006). However, independently of the method of perturbation used, ensemble forecasts derived for a single dynamical forecast system are usually underdispersive or overconfident (i.e. the ensemble spread is smaller than the forecast error), especially at longer lead times (Palmer et al., 2005). This is because initial condition is not the only source of uncertainty in weather and climate prediction, model formulation itself adds a great amount of uncertainty in the forecasting procedure. These uncertainties come from the simplification of the nonlinear equations required to solve them numerically, the limited spatial and temporal resolution of the dynamical forecast systems, which implies that some of the important climate variables are solved through parameterization, and the lack of perfect knowledge of all single aspects of the climate system physics (Palmer, 2000; Curry and Webster, 2011). Therefore, both initial condition and model inadequacy uncertainties must be quantified in order to perform reliable predictions (Palmer et al., 2005; Doblas-Reyes et al., 2009).

Several methods have been proposed to quantify both initial conditions and model inadequacy uncertainties simultaneously (Doblas-Reyes et al., 2009). In one of them, the ensembles of several dynamical forecast systems, developed almost independently by different research groups around the world, are combined to make one forecast. This approach samples model formulation uncertainty by the fact that different institutions resolve the primitive equations differently and use different parameterizations to represent the unresolved physical processes. This method is known as the *multimodel ensemble*. Another method uses a single dynamical forecast system to create a very large ensemble by using multiple combinations of parameterization schemes and parameter values, while avoiding combinations likely to double-count the effect of perturbing a given physical process (Murphy et al., 2004; Meehl et al., 2007). This method is known as the *perturbed-parameter ensemble*. A third approach quantifies model uncertainty by adding stochastic perturbation estimated from parameterized unresolved physical processes to the time derivatives of the primitive equations (Buizza et al., 1999). This approach is known as the *stochastic-physics ensemble*.

Each method has advantages and disadvantages. For instance, the multimodel ensemble is the only method that samples uncertainty from the way the primitive equations are resolved (Doblas-Reyes et al., 2009). On the other hand, the other two methods sample model uncertainty using only a single dynamical forecast system and this gives an important advantage over the multimodel ensemble as it allows one to have greater control over the design of the experiments sampling uncertainty (Murphy et al., 2007). Very few studies have compared the forecast quality of these three representations of model inadequacy. In one of them, Doblas-Reyes et al. (2009) compared the forecast quality of seasonal and annual predictions using these three approaches to account for model inadequacy. They showed that a larger-sized multimodel ensemble outperforms the perturbed-parameter and stochastic-physics ensembles more frequently than not in terms of both deterministic and probabilistic verification measures. However, the multimodel ensemble also outperforms the other methods even if they had the same ensemble size, for predictions at lead times shorter than four months. Figure 2.5 illustrates the average superiority of the multimodel ensemble over the perturbed-parameter and

stochastic-physics ensembles in terms of correlation coefficient, for several seasonal predictions at lead times up to four months. Most significant differences point to a superiority of the multimodel ensemble. In a posterior study, Weisheimer et al. (2011) found that the stochastic-physics ensemble outperforms the multimodel ensemble, especially when predicting precipitation.

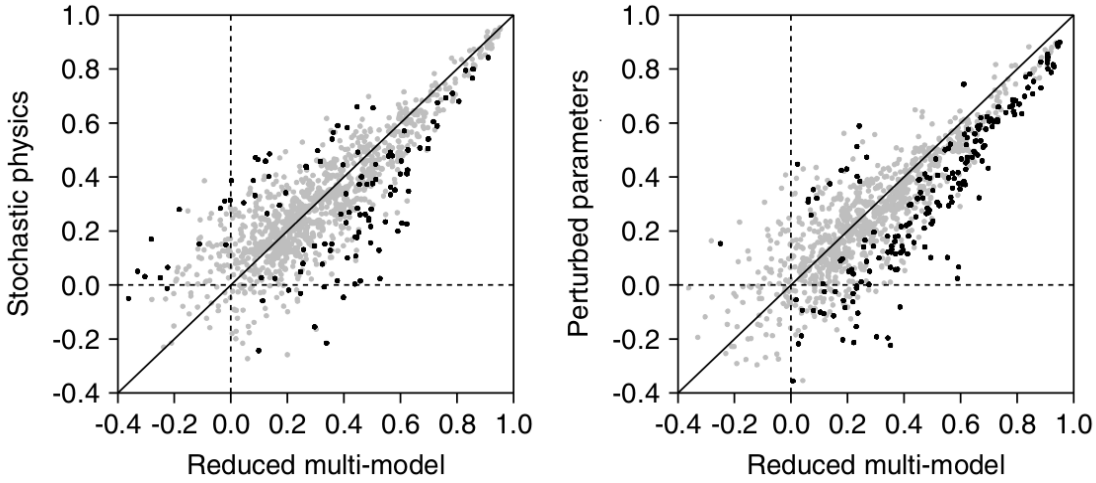

Figure 2.5: Scatter plots comparing the ensemble-mean correlation of different methods used to quantify model uncertainty: (left) stochastic-physics versus reduced multimodel and (right) perturbed-parameter versus reduced multimodel. Each dot shows the ensemble-mean correlation for the seasonal prediction of several climate variables (500 hPa geopotential height, 850 hPa temperature, precipitation, near-surface temperature and mean sea level pressure), two start dates (May and November), four lead times (lead times from zero up to four months), and several regions. Black dots are used for cases where the differences between two forecast systems are statistically significant with 95% confidence. Source: Adapted from Doblas-Reyes et al. (2009).

When applying the multimodel approach, a question that immediately arises is to find the best way to combine the predictions made with the different forecast systems (Knutti, 2010). It has been demonstrated that combining several dynamical forecast systems with equal weights or simple multimodel (SMM) has, on average, improved deterministic and probabilistic forecast quality with respect to the single forecast systems (Palmer et al., 2004; Doblas-Reyes et al., 2005; Hagedorn et al., 2005; Tippett and Barnston, 2008; Wang et al., 2009). Doblas-Reyes et al. (2005) explored several combination methods to merge several dynamical forecast systems setting different weights to each one based on their past performance and using different flavours of multiple linear regression. However, the small sample size typically available in climate prediction produces results with the combination methods that assign unequal weights that are not conclusive or robust, making the SMM a particularly successful benchmark (Doblas-Reyes et al., 2005). Other studies attempted to use more sophisticated combination methods and concluded that it is difficult to improve the SMM forecasts (Kug et al., 2007; Kug et al., 2008; Tippett and Barnston, 2008; DelSole et al., 2012).

In a slightly different framework, Coelho et al. (2004) used a Bayesian method to combine the European Centre for Medium-Range Weather Forecasts (ECMWF) dynamical forecast system with a simple statistical model based on lagged regression to estimate calibrated probabilistic forecasts for the Niño3.4 index. Stephenson et al. (2005) generalized this method to deal with more than one forecast system and more than one variable. They applied this Bayesian method to equatorial Pacific SST grid point predictions produced by seven dynamical forecast systems in the Development of a European Multimodel Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER; Palmer et al., 2004) and showed improved forecast skill compared to individual forecast systems and the SMM.

In this thesis, the multimodel ensemble technique is used to quantify model inadequacy for seasonal predictions of SST over three different ocean basins, two modes of the WAM rainfall variability and near-surface temperature and precipitation over Europe. A comprehensive description of the several variants of the multimodel ensemble is described in Chapter 3.

3. Data and methods

3.1. Observations

Two observational SST datasets have been used: the Hadley Center's Global Sea-Ice Coverage and Sea Surface Temperature version 1.1 (HadISSTv1.1; Rayner et al., 2003) and the Extended Reconstructed Sea Surface Temperature analysis version v3b (ERSSTv3b, Smith et al., 2008). HadISSTv1.1 contains a set of monthly fields of global SST and sea ice concentration on a 1º latitude and longitude grid from 1871 onwards while ERSSTv3b is generated using in situ SST data and improved statistical methods that allow stable reconstruction using sparse data. ERSSTv3b has a 2º resolution and covers the period from January 1854 onwards. HadISSTv1.1 is used in Chapter 4 to build linear regression models and in the forecast quality assessment. ERSSTv3b is used as predictor to estimate the predictors to build the statistical models to predict the WAM precipitation regimes in Chapter 5 and spatial-fields of precipitation and near-surface temperature over Europe in Chapter 6.

Two observational precipitation datasets have been used: the version 2.2 of the Global Precipitation Climatology Project (GPCP) monthly satellite-gauge combination (Huffman and Bolvin, 2013) and the Global Precipitation Climatology Center (GPCC) version 6.0 monthly gridded gauge analysis derived from quality controlled station data (Schneider et al., 2011). The GPCP dataset is available at a 2.5º resolution for the period from 1979 onwards and covers land and ocean. On the other hand, the 1º resolution GPCC dataset is available only over land for the period from 1901 onwards. The GPCP dataset is used for the forecast quality assessment and the GPCC dataset is used to estimate the linear regression coefficients of the statistical models applied in Chapters 5 and 6.

Two observational near-surface temperature datasets have been used: ERA-Interim (Dee et al., 2011) and the Global Historical Climatology Network monthly version 2 (GHCNv2; Fan and van den Dool, 2008). ERA-Interim is the latest global atmospheric reanalysis produced by the ECMWF. It is presented as a gridded dataset at approximately 0.7º horizontal resolution and covers the period from January 1979 onwards. GHCNv2 is an in-situ observation-based global land monthly mean near-surface temperature at a 2.5º x 2.5º horizontal resolution for the period 1948 onwards.

3.2. Dynamical forecast systems

Several operational and quasi-operational dynamical forecast systems are used in this thesis, among them two from the European Seasonal to Interannual Prediction (EUROSIP) initiative and six from the North America Multimodel Ensemble (NMME) project. Besides, a simple statistical model is used as a benchmark for comparison with the dynamical forecast systems. All forecast systems are described below.

The atmospheric component of the ECMWF climate forecast system 4 (S4) is the cycle 36r4 of the ECMWF Integrated Forecast System (IFS) (Molteni et al., 2011; Kim et al., 2012). It has a horizontal resolution of about 80 km and 91 vertical levels, extending up to about 0.01 hPa. The ocean component of S4, the Nucleus for European Modelling of the Ocean (NEMO) version 3.0, has a horizontal resolution of about 1º with equatorial refinement and 42 vertical levels, 18 of which are in the upper 200 m. S4's hindcasts have 15 ensemble members, except the ones started in February, May, August and November, which have 51 members. All ensemble members starts in burst mode on the first day of every month at 0 UTC and the simulations are seven-month long and cover the period 1981-2011.

The National Center for Environmental Prediction (NCEP) climate forecast system version 2 (CFSv2) uses the Global Forecast System (GFS), with horizontal resolution of about 100 km and 64 vertical levels, as its atmospheric component (Yuan et al., 2011; Kim et al., 2012; Kirtman et al., 2014; Saha et al., 2014). Its ocean component is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 4 (MOMv4) and it has maximum horizontal resolution of 0.25º within 10º of the equator and 0.5º poleward and 40 vertical levels. CFSv2 hindcasts have 24 ensemble members, except those starting in November, which have 28 members. CFSv2´s hindcasts are initialized in different days and times, being the ones initialized after the day 7 used as the lead time zero ensemble members of the next month. For example, the ensemble members for the target month of February at lead time zero have start dates in January 11th, 16th, 21st, 26th, 31st, and the February 5th (at the synoptic times 00, 06, 12 and 18 UTC) of the same year. The simulations are ten-month long and cover the period 1982-2011.

The Météo-France seasonal forecast system version 3 (MF3) uses the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) version 4 as its atmospheric component (Alessandri et al., 2011). It has a horizontal resolution of about 300 km and 91 vertical levels, reaching high into the stratosphere. Its ocean component is the global version of the Océan PArallélisé (OPA) model version 8.2 with a horizontal resolution of about 2º and 31 vertical levels. MF3´s hindcasts have 11 ensemble members, all starting in burst mode on the first day of every month at 0 UTC. The simulations are seven-month long and cover the period 1981-2011.

The Community Climate System Model version 3 (CCSM3) uses the Community Atmosphere Model (CAM) version 3, with horizontal resolution of approximately 150 km and composed of 26 vertical levels (Yoshikatsu et al., 2008; Kirtman and Min, 2009; Kirtman et al., 2014), as its atmospheric component. The Parallel Ocean Program (POP) with 1º horizontal resolution and 40 vertical levels is the ocean component (Yoshikatsu et al., 2008). CCSM3´s hindcasts have 6 ensemble members, all starting in burst mode on the first day of every month at 0 UTC. The simulations are twelve-month long and cover the period 1982-2011.

The GFDL uses the GFDL Atmospheric Model with horizontal resolution of roughly 200 km and 24 vertical levels as its atmospheric component and the MOMv4 with maximum horizontal resolution of about 0.3º near the Equator (1º elsewhere) and 50 vertical levels as it ocean component (Zhang et al., 2007; Kirtman et al., 2014). Its hindcasts have 10 ensemble members, all starting in burst mode on the first day of every month at 0 UTC. The simulations are twelve-month long and cover the period 1982-2011.

The International Research Institute for Climate and Society - European Center Hamburg Model (IRI-ECHAM) anomaly and IRI-ECHAM direct use the coupled forecast system described in DeWitt (2005) with some updated parameterizations. The atmospheric component is the ECHAM version 4.5 with horizontal resolution of about 300 km and 19 vertical levels. The ocean component is the MOM version 3 (MOMv3) with zonal resolution of 1.5º and meridional resolution of 0.5º between 10ºS and 10ºN, gradually increasing to 1.5º, keeping constant at this value north of 30ºN and south of 30ºS. There are 25 vertical layers and 17 layers in the upper 450 m. Both forecast systems produce hindcasts with 12 ensemble members, all of them starting in burst mode on the first day of every month at 0 UTC and are nine-month long. They cover the common period 1982-2011. The difference between the two versions of the IRI system is that the IRI-ECHAM direct employs direct coupling while the IRI-ECHAM anomaly employs anomaly coupling.

The Canadian Meteorological Center seasonal forecast system version 2 (CMC2) uses the Canadian Center for Climate Modelling and Analysis (CCCam) atmospheric circulation model version 4 (CanAM4) as its atmospheric component (Merryfield et al., 2013; Kirtman et al., 2014). CanAM4 has a horizontal resolution of about 200 km and 35 vertical levels. The ocean component is the CCCma ocean model version 4 (CanOM4) with horizontal resolution of approximately 100 km and 40 vertical levels. CMC2´s hindcasts have 10 ensemble members, all starting in burst mode on the first day of every month at 0 UTC. The hindcasts are twelve-month long and cover the period 1981-2011.

The National Aeronautics and Space Administration (NASA) forecast system has the Goddard Earth Observing System version 5 (GEOS5) model ACGM as its atmospheric component (Vernieres et al., 2012). GEOS5 has a horizontal resolution of about 200 km and 72 vertical levels. MOMv4 with maximum horizontal resolution of about 0.25º near the Equator (1º elsewhere) and 40 vertical levels is its ocean component. NASA´s hindcasts have 11 ensemble members, four of which start every five days while the other members have other perturbation methods applied on the day closest to the beginning of the month. The hindcasts are nine-month long and cover the period 1981-2010.

3.3. Statistical model

The statistical models used in Chapters 4, 5 and 6 are based on simple linear regression where the relationship between a vector of predictands and a vector of predictors is

estimated using the least squares method. The equations used to estimate the statistical model are described below. The vectors of predictands can be written as:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

where y_i is the predictand at the ith target year and N is the number of target years. For example, y_i could be an observed Niño3.4 SST index in June 2000. The vector of predictors of the statistical model can be written as:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

where x_i is the predictor at the ith target year. In this thesis, the statistical model uses observations as predictors. For example, x_i could be an observed Niño3.4 SST index in May 2000. The prediction of this example would be a one-month forecast with lead time zero because it uses the latest observations prior to the target month. The simple linear regression can be expressed in matrix form as follows:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a} + \boldsymbol{\varepsilon} \tag{3.1}$$

where

$$\boldsymbol{a} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

and

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$$

where a_0 and a_1 are the least-squares estimates of the intercept and the slope parameters, respectively, and $\boldsymbol{\varepsilon}$ is the vector of residuals. Note that $\boldsymbol{X}$ is an extension of the vector of predictors. The least-squares estimate of $\boldsymbol{a}$ is obtained in cross-validation mode by minimizing the sum of the squared error (SSE), $SSE = \left(\boldsymbol{y}^i - \boldsymbol{X}^i\boldsymbol{a}^i\right)^T\left(\boldsymbol{y}^i - \boldsymbol{X}^i\boldsymbol{a}^i\right)$, and has the following standard solution:

$$\boldsymbol{a}^i = \left[\left(\boldsymbol{X}^i\right)^T\boldsymbol{X}^i\right]^{-1}\left[\boldsymbol{X}^i\right]^T\boldsymbol{y}^i \tag{3.2}$$

where the superscript i indicates that all years but the ith target year are included in the regression analysis because the estimates are obtained in one-year-out cross-validation mode. This method is applied to estimate the statistical model presented in Chapter 4. The statistical model is also estimated in retroactive mode where only years prior to the target year are used in the estimation of the regression coefficients, as in an operational context (Mason and Mimmack, 2002; Mason and Baddour, 2008). In this case, the superscript i indicates that only the years prior to the ith target year are included in the regression

analysis. In both cases, cross-validation and retrospective approaches, the statistical model prediction at the ith target year is estimated as:

$$\hat{y}_i = x_i a_1^i + a_0^i \tag{3.3}$$

Due to the chaotic nature of the climate system, seasonal forecasts should be formulated in a probabilistic fashion (e.g. Doblas-Reyes et al., 2013a). Therefore, the statistical model must be communicated with a proper quantification of the forecast uncertainty. The first step to quantify the forecast uncertainty in the statistical model is to estimate the predicted variance, which is a measure of spread. If the residuals are assumed to follow a Gaussian distribution and that a large dataset is used in the regression analysis, then the unbiased estimate of the predicted variance would be equal to the mean of the squared regression residuals:

$$\hat{\sigma}_{0i}^2 = \frac{1}{n-2}\sum_{j=1}^{n} \varepsilon_j^2 \tag{3.4}$$

where n is the number of training years (e.g., $n = N - 1$ when the statistical model is estimated in one-year-out cross-validation mode), the factor $n - 2$ appears because two regression coefficients (a_0^i and a_1^i) are estimated, and ε_j is the estimated residual at the jth training year. However, because of the small sample size usually available to perform regression analysis in seasonal forecast, $\hat{\sigma}_{0i}$ usually underestimates the forecast uncertainty (Wilks, 2006). To minimize this issue, two additional terms must be added in the estimation of the predicted variance so that the predicted variance can be rewritten as:

$$\hat{\sigma}_i^2 = \hat{\sigma}_{0i}^2 \left[1 + \frac{1}{n} + \frac{\left(x_i - \bar{x}^i\right)^2}{n\gamma^{2,i}}\right] \tag{3.5}$$

where

$$\gamma^{2,i} = \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \bar{x}^i\right)^2 \tag{3.6}$$

where $\bar{x}^i$ and $\gamma^{2,i}$ are the statistical model´s predictor mean and variance over the training period for the ith target year. Note in the third term of the equation (3.5) that the forecast will be more uncertain when the predictor x_i is far from the center of its climatological distribution over the training period. On the other hand, this uncertainty would become negligible when a very large sample size is used to estimate the regression coefficients (Wilks, 2006). For an infinite sample size $\hat{\sigma}_i^2$ would equal to $\hat{\sigma}_{0i}^2$. Assuming the predicted PDF follows a Gaussian distribution, then the 95% prediction interval is expected to be approximately bounded by $\hat{y}_i \pm 1.96\hat{\sigma}_i$. When this assumption fails, the prediction interval tends to be either too wide or too narrow.

3.3.1. SST forecasts

In Chapter 4, the statistical model is used to predict monthly anomalies of three SST indices over different tropical regions: the Niño3.4 (170ºW - 120ºW, 5ºS - 5ºN; Trenberth, 1997), the Subtropical Northern Atlantic (SNA; 55ºW - 15ºW, 5ºN - 25ºN; Enfield et al., 1999), and the Western Tropical Indian (WTI; 50ºE - 70ºE, 10ºS - 10ºN; Wang et al. 2009). The statistical model assumes the predictor and the predictand are the same index, but at different target months. For example, the Niño3.4 SST index in May 2000 could be used as predictor to predict the Niño3.4 SST index in June 2000 for a forecast at lead time 0, as it uses the latest observed SST index as predictor. Forecasts are performed for all months of the year and the seven forecast times available in common for both S4 and MF3.

For each target month and lead time pair, the statistical model is trained in two different ways. On one hand, the one-year-out cross-validation method is applied using the period 1951-2010. On the other hand, the regression coefficients are estimated in retroactive mode having the period 1951-1981 as the first training period for the forecasts performed for the target years 1982-2010, extending the training period by one year at a time as in an operational context (Mason and Mimmack, 2002; Mason and Baddour, 2008). As for the forecast quality assessments of the dynamical forecast systems, verification statistics are computed for the target period 1982-2010. The HadISSTv1.1 is used both in the estimation of the regression coefficients and in the forecast quality assessment.

A brief description of the differences in skill between the statistical model predictions in retroactive and cross-validation modes is provided in Appendix A. The statistical model developed in retroactive mode is used in Chapter 4.

3.3.2. WAM rainfall regimes

In Chapter 5, the methodology used to estimate the WAM rainfall regimes is aimed to take into account the seasonal evolution of the rainfall within a rainy season and its interannual variability simultaneously. In this technique, monthly rainfall is averaged over the 10ºW-10ºE African Monsoon Multidisplinary Analysis transect (Hourdin et al., 2010; Losada et al., 2010; Roehrig et al., 2013). Averaging rainfall zonally allows taking into account two relevant features of the WAM variability: the latitudinal migration and the seasonal distribution of the summer rainfall (Hourdin et al., 2010). The latitudinal range of the study extends from the Equator to 20ºN and the period between June and October of each year. The southernmost limit is intended to capture the inland penetration of monsoonal rainfall over the Guinean region, while the northernmost limit tries to capture the Sahelian rainfall, which usually reaches 18ºN in the observations. The period from June to October represents one month prior to and one month after the July, August and September (JAS) period, a time when most rainfall takes place over the WAM region (Sultan and Janicot, 2003).

With the dynamical forecast systems analyzed in this thesis, the longest forecast time that can be considered is seven months, the longest forecast time of both S4 and MF3. With seven forecast months and covering the period June to October, three start dates can be considered to estimate the intraseasonal evolution of the WAM rainfall: June (lead 0), May (lead 1 month) and April (lead 2 months). Most users, such as farmers, request receiving information about seasonal rainfall about 1-2 months before the climatological monsoon onset in July, that is, the information should be available in late April or early May (Ingram et al., 2002). For this reason, most figures displayed in Chapter 5 are for lead time 1, that is, a prediction starting in the first of May or late April (as is the case for the CFSv2).

Principal component analysis (PCA; Wilks, 2006) of the covariance matrix is performed upon the observed and predicted zonally averaged rainfall to estimate the leading modes of WAM rainfall variability. The three-dimensional data matrix used to estimate the covariance matrix contains the longitudes, the months from June to October and the number of years, which will identify the modes of interannual variability taking into account at the same time the seasonal variability. For the hindcasts, the third dimension is the number of ensemble members times the number of years. The anomalies, estimated for both the observations and the dynamical forecast systems prior to applying the PCA, were computed using three-year-out cross-validation to avoid artificial skill in the forecast quality assessment (Mason and Baddour, 2008). The leading modes of the WAM rainfall variability are described as a set of spatial patterns (empirical orthogonal functions, EOFs) and associated standardized time series (PCs) that are associated to specific modes of variability. PCA is performed upon the observations and each forecast system and lead time separately to take into account the fact that the hindcasts might represent the variability in a way different to the observations, while this representation also depends on the lead time (Doblas-Reyes et al., 2003; Philippon et al., 2010).

Statistical models based on simple linear regression are used to predict the WAM rainfall regimes. The regression coefficients are estimated in retroactive mode having the first training period the years 1951-1981, extending it by one year at a time for the entire target period 1982-2011. The selection of two periods has different motivations. While the year 1982 is the first year available for most hindcasts, 1951 is the year from which a large number of stations are used in the GPCC dataset, and therefore, making this period a more trustworthy period for this dataset (Schneider et al., 2011). The PCA is performed on the GPCC and GPCP to assess the uncertainty associated with the observations in the estimation of the WAM rainfall regimes and to use an independent dataset for the estimation of the statistical model from the one used in the forecast quality assessment. The GPCP dataset is also used with a mask over the ocean for comparison with the GPCC data, which has values only over land. Climatologies and the PCA for the observed rainfall are hence computed with four different samples: GPCP 1982-2011, GPCP land-only 1982-2011, GPCC 1951-2011 and GPCC 1982-2011. The GPCC and ERSSTv3b datasets are used to build the statistical model while the GPCP dataset is used for the validation of all forecast system.

Chapter 5 focuses on the two regimes that explain most of the rainfall variability in the WAM region. The statistical model used to predict these regimes use SST indices as predictors. Three SST indices are considered: the Atlantic 3 (Atl3; SST averaged over 20ºW-0ºE, 3ºS-3ºN; Zebiak, 1993), the Atlantic Multidecadal Oscillation (AMO; SST anomalies averaged over 80ºW-0ºW and 0º-60ºN minus global SST anomalies over 60ºS-60ºN; Trenberth and Shea, 2006) and the Niño3.4 (SST anomalies averaged over 170ºW-120ºW; 5ºS-5ºN; Trenberth, 1997). The choice of the best predictor for each WAM rainfall regime is given in the Appendix B.

3.3.3. Near-surface temperature and precipitation over Europe

In Chapter 6, the statistical model is used to predict spatial-fields of monthly near-surface temperature and precipitation over Europe, in boreal summer and winter. The Niño3.4, SNA and AMO SST indices as well as the predictand variable itself at previous months are used as predictors. The statistical model is estimated in retroactive mode and the relation between the predictor and the predictand is estimated for each grid point independently. The first training period is 1951-1981, extending it by one year at a time, to predict the target period 1982-2010. ERSSTv3b is used to estimate the SST indices. The GPCC and the GHCNv2 are used to train the statistical models while the GPCP and the ERA-Interim are used in the forecast quality assessment. Because the observations and the dynamical forecast systems have different spatial resolution, they have been interpolated into the 2.5º grid resolution prior to the estimation of the regression coefficients of the statistical model.

3.4. Combination methods

The dynamical forecast systems described previously are combined using different combination methods. This section presents a quantitative description of the combination methods used in Chapters 4, 5 and 6. The spatial-fields of the dynamical forecast systems can be organized in matrix form as follows:

$$\widehat{\boldsymbol{F}}_{i,k} = \begin{bmatrix} \hat{e}_{1,1} & \hat{e}_{1,2} & \cdots & \hat{e}_{1,J^*} \\ \hat{e}_{2,1} & \hat{e}_{2,2} & \cdots & \hat{e}_{2,J^*} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{e}_{I^*,1} & \hat{e}_{I^*,2} & \cdots & \hat{e}_{I^*,J^*} \end{bmatrix}$$

where $\hat{e}$ is a grid-point ensemble member of the kth forecast system at the ith target year for a given variable, target period and lead time. I^* and J^* are the number of latitudes and longitude points in the forecast system´s horizontal resolution. When the combination is performed to predict spatial-fields of climate variables, as the ones in Chapter 6, the forecast systems are first interpolated into the 2.5º grid. In this case, after computing the ensemble mean for each grid-point in a forecast system and first interpolated into the 2.5º grid, the matrix $\widehat{\boldsymbol{F}}$ can be rewritten as:

$$\widehat{\boldsymbol{S}}_{i,k} = \begin{bmatrix} \hat{\hat{e}}_{1,1} & \hat{\hat{e}}_{1,2} & \cdots & \hat{\hat{e}}_{1,J} \\ \hat{\hat{e}}_{2,1} & \hat{\hat{e}}_{2,2} & \cdots & \hat{\hat{e}}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\hat{e}}_{I,1} & \hat{\hat{e}}_{I,2} & \cdots & \hat{\hat{e}}_{I,J} \end{bmatrix}$$

where $\hat{\hat{e}}$ is the grid-point ensemble mean of the kth forecast system at the ith target year for a given variable, target period and lead time. I and J are the number of latitudes and longitude points in the 2.5º horizontal resolution. For example, $\widehat{\boldsymbol{S}}_{i,k}$ could be the spatial-field of precipitation predicted by S4 in May 2000 at lead time 0. All observations described in the Section 3.1 are also interpolated into the 2.5º horizontal resolution. The interpolated observations can be written generically in matrix form as:

$$\boldsymbol{Y}_i = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,J} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ y_{I,1} & y_{I,2} & \cdots & y_{I,J} \end{bmatrix}$$

where y is the observed climate variable at the ith target year for a given target period. For example, it could be the spatial-field of precipitation in May 2000.

In Chapters 4 and 5, the combinations are performed for univariate predictands. The matrix $\widehat{\boldsymbol{F}}$ is used to compute the SST indices used in Chapter 4 as well as the WAM rainfall regimes in Chapter 5. In these cases, when only univariate variables are used, the matrix representing the dynamical forecast systems can be rewritten as:

$$\widehat{\boldsymbol{M}} = \begin{bmatrix} \widehat{m}_{1,1} & \widehat{m}_{1,2} & \cdots & \widehat{m}_{1,M} \\ \widehat{m}_{2,1} & \widehat{m}_{2,2} & \cdots & \widehat{m}_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{m}_{N,1} & \widehat{m}_{N,2} & \cdots & \widehat{m}_{N,M} \end{bmatrix}$$

where $\widehat{m}_{i,k}$ is the ensemble mean of a forecast system, N is the number of target years and M is the number of forecast systems. Note that $\widehat{m}_{i,k}$ is estimated in the forecast system´s original resolution. For example, $\widehat{m}_{i,k}$ could be an S4´s ensemble mean prediction of the Niño3.4 SST index for the target period of June 2000 with lead time one. Note that $\boldsymbol{y}$, $\boldsymbol{Y}$ and $\widehat{\boldsymbol{M}}$ refer to anomalies computed using the three-year-out cross-validation method, that is, for each case the mean was computed using all years, but the $i-1$th, ith and $i-1$th target years. With the exception of the SMM, where no calibration is applied, the combination of the forecast systems is also performed in three-year-out cross-validation mode.

3.4.1. Simple multimodel

The SMM is defined by the simple average of the ensemble mean of several forecast systems, such as:

$$\hat{y}_i^{SMM} = \frac{1}{M}\sum_{k=1}^{M} \widehat{m}_{i,k} \tag{3.7}$$

where $\hat{y}_i^{SMM}$ is the ensemble mean of the SMM at the ith target year. The ensemble mean of each dynamical forecast system is computed prior to the combination to avoid giving more weight to the forecast systems that has a larger number of ensembles. Note that the equation (3.7) is applied to each target period, variable, lead time and region independently.

3.4.2. Multiple linear regression

A multiple linear regression (MLR) of the observations on the anomaly values of the forecast systems was performed to estimate the linear combination of the different forecast systems. The MLR combination can be expressed in matrix form as follows:

$$\boldsymbol{y} = \boldsymbol{E}\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{3.8}$$

where

$$\boldsymbol{E} = \begin{bmatrix} 1 & \hat{m}_{1,1} & \hat{m}_{1,2} & \cdots & \hat{m}_{1,M} \\ 1 & \hat{m}_{2,1} & \hat{m}_{2,2} & \cdots & \hat{m}_{2,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{m}_{n,1} & \hat{m}_{n,2} & \cdots & \hat{m}_{n,M} \end{bmatrix}$$

and

$$\boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_M \end{bmatrix}$$

where b_0 and b_k are the least-squares estimates of the intercept and the slope parameters of the kth forecast system and $\boldsymbol{\varepsilon}$ is the vector of residuals. Note that $\boldsymbol{E}$ is an extension of the matrix of predictors $\boldsymbol{M}$ and n is the number of training years. The least-squares estimate of $\boldsymbol{b}$ is obtained using the three-year-out cross-validation method by minimizing the SSE using the equation 3.2, replacing $\boldsymbol{X}$ for $\boldsymbol{E}$ and $\boldsymbol{a}$ for $\boldsymbol{b}$. In this case, the superscript i indicates that all years, but the $i-1$th, ith and $i+1$th target years were included in the regression analysis. Thus, the predicted value for the MLR combination at the ith target year is then estimated as:

$$\hat{y}_i^{MLR} = \sum_{k=1}^{M} \hat{m}_{i,k} b_k^i + b_0^i \tag{3.9}$$

The covariance matrix of the forecast system´s predictions, necessary to estimate the predicted standard deviation, was computed as follows:

$$\boldsymbol{S}_{mm}^{2,i} = \frac{1}{n}\left(\boldsymbol{M}^{i}\right)^{T}\boldsymbol{M}^{i} = \begin{bmatrix} s_{1,1}^{2,i} & s_{1,2}^{2,i} & \cdots & s_{1,M}^{2,i} \\ s_{2,1}^{2,i} & s_{2,2}^{2,i} & \cdots & s_{2,M}^{2,i} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M,1}^{2,i} & s_{M,2}^{2,i} & \cdots & s_{M,M}^{2,i} \end{bmatrix} \tag{3.10}$$

where the superscript i indicates that all, but the $i-1$th, ith and $i+1$th target years are included in the computation of the variances-covariances, and $s_{k,k}^{2,i}$ and $s_{k,l}^{2,i}$ are the variances and covariances given by:

$$s_{k,k}^{2,i} = \frac{1}{n}\sum_{j=1}^{n}\left(\hat{m}_{j,j}\right)^{2} \qquad \forall j \neq i-1, i, i+1 \tag{3.11}$$

$$s_{k,l}^{2,i} = \frac{1}{n}\sum_{j=1}^{n}\left(\hat{m}_{j,j}\hat{m}_{j,j^{*}}\right) \qquad \forall j \neq i-1, i, i+1 \tag{3.12}$$

where $n = N - 3$ is the number of training years and l is a forecast system, such as $j \neq j^{*}$.

The predicted forecast uncertainty for each target year was computed as in Doblas-Reyes et al. (2005):

$$\hat{s}_{i}^{MLR} = s_{0}^{i}\sqrt{1 + \frac{1}{n}\hat{\boldsymbol{m}}_{i}\left(\boldsymbol{S}_{mm}^{2,i}\right)^{-1}\hat{\boldsymbol{m}}_{i}^{T}} \tag{3.13}$$

where s_{0}^{i} is the standard deviation of the regression residuals and $\hat{\boldsymbol{m}}_{i}$ is the vector of predictors at the ith target year, such as $\hat{\boldsymbol{m}}_{i} = \left[\hat{m}_{i,1}\ \hat{m}_{i,2}\ \cdots\ \hat{m}_{i,M}\right]$.

3.4.3. Principal component multiple linear regression

A PCA is performed on $\boldsymbol{M}^{i}$ aiming at finding a new set of predictors that are uncorrelated from each other. This new set of predictors is used to estimate the PC1-regression (PC1) and the PCA-regression (PCA-regression) combinations. The aim of using PCA is to avoid introducing a large uncertainty in the estimated linear regression coefficients due to colinearity (Doblas-Reyes et al., 2005).

The eigenvalue decomposition of the covariance matrix $\boldsymbol{S}_{mm}^{i}$ can be written as

$$\boldsymbol{S}_{mm}^{i}\boldsymbol{F}^{i} = \boldsymbol{\lambda}^{i}\boldsymbol{F}^{i} \tag{3.14}$$

where $\boldsymbol{F}^{i}$ and $\boldsymbol{\lambda}^{i}$ are the eigenvectors and eigenvalues of $\boldsymbol{S}_{mm}^{i}$, respectively.

The PC, $\boldsymbol{P}^i$ are given by

$$\boldsymbol{P}^i = \boldsymbol{M}^i \cdot \boldsymbol{F}^i \tag{3.15}$$

The PCs for the ith target year $\boldsymbol{p}_i$ is computed by multiplying the matrix of eigenvectors $\boldsymbol{F}^i$ by the vector of predictors $\widehat{\boldsymbol{m}}_i = \left[\widehat{m}_{i,1}\ \widehat{m}_{i,2}\ \cdots\ \widehat{m}_{i,M}\right]$ at the ith target year.

In the PC1 combination, the PC that explained the largest variance is used to compose $\boldsymbol{E}$ and the steps (3.8) to (3.13) are performed to estimate the predicted value $\hat{y}_i^{PC1}$ and the predicted standard deviation $\hat{s}_i^{PC1}$ of the PC1 combination. In the PCA-regression combination, all PCs are used to compose $\boldsymbol{E}$ and the predicted value $\hat{y}_i^{PCA}$ and the predicted standard deviation $\hat{s}_i^{PCA}$ of the PCA-regression combination are estimated.

3.4.4. Forecast assimilation

The forecast assimilation (FA) is a Bayesian approach that combines the dynamical forecast system predictions with prior historical information to produce calibrated probabilistic forecasts (Stephenson et al., 2005). It can be expressed as:

$$y_i|m_i = N(y_i, s_i) \tag{3.16}$$

The predicted mean $\hat{y}_i$ and the predicted standard deviation $\hat{s}_i$ can be written as follows:

$$\hat{y}_i = y_b^i + \boldsymbol{L}^i\left[\widehat{\boldsymbol{m}}_i - \boldsymbol{G}^i\left(y_b^i - y_0^i\right)\right] \tag{3.17}$$

$$\hat{s}_i^2 = \left[\left(\boldsymbol{G}^i\right)^T\left(\boldsymbol{S}^{2,i}\right)^{-1}\boldsymbol{G}^i + \left(\boldsymbol{C}^i\right)^{-1}\right]^{-1} \tag{3.18}$$

where the $\boldsymbol{L}^i = \boldsymbol{C}^i\left(\boldsymbol{G}^i\right)^T\left[\boldsymbol{G}^i\boldsymbol{C}^i\left(\boldsymbol{G}^i\right)^T + \boldsymbol{S}^{2,i}\right]^{-1}$ is the gain/weight matrix. The slope $\boldsymbol{G}^i$, the intercept y_0^i and the prediction error covariance $\boldsymbol{S}^{2,i}$ matrices are estimated using the least-squares estimation of the regression of the forecast systems on the observations. They are given by:

$$\boldsymbol{G}^i = \boldsymbol{S}_{my}^{2,i}\left(\boldsymbol{S}_{yy}^{2,i}\right)^{-1} \tag{3.19}$$

$$y_0^i = -\left[\bar{\boldsymbol{m}}^i - \bar{y}^i\left(\boldsymbol{G}^i\right)^T\right]\boldsymbol{G}^i\left[\left(\boldsymbol{G}^i\right)^T\boldsymbol{G}^i\right]^{-1} \tag{3.20}$$

$$\boldsymbol{S}^{2,i} = \boldsymbol{S}_{mm}^{2,i}\left(\boldsymbol{S}_{yy}^{2,i}\right)^{-1}\left(\boldsymbol{S}_{my}^{2,i}\right)^T \tag{3.21}$$

where $\boldsymbol{S}_{yy}^{2,i}$ is the covariance matrix of the observations, and $\boldsymbol{S}_{my}^{2,i}$ is the cross-covariance matrix. $\bar{\boldsymbol{m}}^i$ and $\bar{y}^i$ are defined as:

$$\bar{\boldsymbol{m}}^i = \frac{1}{n}\sum_{j=1}^{n} \hat{m}_{j,*} \qquad \forall j \neq i-1, i, i+1 \tag{3.22}$$

$$\bar{y}^i = \frac{1}{n}\sum_{j=1}^{n} y_j \qquad \forall j \neq i-1, i, i+1 \tag{3.23}$$

The reader will note that the computations are performed in three-years-out cross-validation mode. The mean (y_b^i) and covariance (C^i) matrices of the normally-distributed prior are computed in two different ways:

- In one case, the expected value and the predicted standard deviation from the statistical model predictions are used as y_b^i and C^i, respectively. This combination is called the FA-statistical (FAS).
- In a second case, the prior distribution was estimated using the climatological information, such as:

$y_b^i = \bar{y}^i$

$C^i = \boldsymbol{S}_{yy}^{2,i}$

This last combination is referred to as the FA-climatology (FAC).

3.4.5. Combination of spatial-fields

Until now, it has been described the methods used to combine time series (i.e., univariate statistics). In Chapter 4, these methods are used to combine predictions of three SST indices and, in Chapter 5, some of the methods are used to predict two WAM rainfall regimes. The methods used to combine the spatial-fields of precipitation and near-surface temperature prediction over Europe in winter and summer are described below. The area of study falls between 20ºN and 75ºN of latitude and between 30ºW and 60ºE of longitude.

The equation (3.7) can be rewritten in matrix form to estimate the spatial-field of the SMM predictions, such as:

$$\widehat{\boldsymbol{Y}}_i^{SMM} = \frac{1}{M}\sum_{k=1}^{M} \widehat{\boldsymbol{S}}_{i,k} \tag{3.24}$$

where $\widehat{\boldsymbol{S}}_{i,k}$ is a spatial-field of the kth forecast systems at the ith target year for a given variable, target period and lead time.

The FA combination is also applied into spatial-field forecasts. First, the matrix $\boldsymbol{Y}$ is reorganized to place the grid points in the rows and the number of training years in the columns, such as:

$$\boldsymbol{Y}^{*} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,Q} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,Q} \end{bmatrix}$$

where n is the number of training years and Q is the number of latitudes times the number of longitudes. Similarly, the matrix $\widehat{\boldsymbol{S}}_{i,k}$ is reorganized to place all grid points and forecast systems in the rows, such as:

$$\boldsymbol{A} = \begin{bmatrix} \hat{\hat{e}}_{1,1} & \hat{\hat{e}}_{1,2} & \cdots & \hat{\hat{e}}_{1,P} \\ \hat{\hat{e}}_{2,1} & \hat{\hat{e}}_{2,2} & \cdots & \hat{\hat{e}}_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\hat{e}}_{n,1} & \hat{\hat{e}}_{n,2} & \cdots & \hat{\hat{e}}_{n,P} \end{bmatrix}$$

where P is the number of grid points in the 2.5º grid (Q) times the number of forecast systems (M). Note that all observations and forecast systems have the same number of grid points after the interpolation. Due to the large dimensionality of gridded data compared to the number of independent forecasts, especially for multimodel predictions (i.e., $P \gg n$), and the strong dependency between values at adjacent grid-points, dimension reduction becomes necessary (Stephenson et al., 2005). Maximum covariance analysis (MCA) is used to extract the leading co-varying modes from the forecast systems predictions and the observations. The cross-covariance matrix can be decomposed into the product of three matrix, such as:

$$\boldsymbol{Y}^{*T}\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{T} \tag{3.25}$$

where the columns of $\boldsymbol{U}_{Q,Q}$ are the orthonormal eigenvectors of $\left(\boldsymbol{Y}^{*T}\boldsymbol{A}\right)\left(\boldsymbol{Y}^{*T}\boldsymbol{A}\right)^{T}$, the columns of $\boldsymbol{V}_{P,P}$ are the orthonormal eigenvectors of $\left(\boldsymbol{Y}^{*T}\boldsymbol{A}\right)^{T}\left(\boldsymbol{Y}^{*T}\boldsymbol{A}\right)$ and $\boldsymbol{D}_{Q,P}$ is a diagonal matrix containing the square roots of the eigenvalues of $\boldsymbol{U}$ or $\boldsymbol{V}$. The PCs associated with the eigenvectors can be written as:

$$\boldsymbol{Z} = \boldsymbol{Y}^{*}\boldsymbol{U}_{Q,nm} \tag{3.26}$$

$$\boldsymbol{W} = \boldsymbol{A}\boldsymbol{V}_{P,nm} \tag{3.27}$$

where $\boldsymbol{Z}$ and $\boldsymbol{W}$ are the left (observations) and right (predictions) expansion coefficients and nm is the number of modes retained. Finally, the equations (3.19) to (3.21) are rewritten to solve the likelihood parameters of the predictions of spatial-fields of climate variables:

$$\boldsymbol{\mathcal{G}}^{i} = \boldsymbol{\mathcal{S}}_{ZW}^{2,i}\left(\boldsymbol{\mathcal{S}}_{ZZ}^{2,i}\right)^{-1} \tag{3.28}$$

$$\boldsymbol{z}_{0}^{i} = -\left[\bar{\boldsymbol{w}}^{i} - \bar{\boldsymbol{z}}^{i}\left(\boldsymbol{\mathcal{G}}^{i}\right)^{T}\right]\boldsymbol{\mathcal{G}}^{i}\left[\left(\boldsymbol{\mathcal{G}}^{i}\right)^{T}\boldsymbol{\mathcal{G}}^{i}\right]^{-1} \tag{3.29}$$

$$\boldsymbol{S}^{2,i} = \boldsymbol{S}_{WW}^{2,i}\left(\boldsymbol{S}_{ZZ}^{2,i}\right)^{-1}\left(\boldsymbol{S}_{ZW}^{2,i}\right)^{T} \tag{3.30}$$

where $\overline{\boldsymbol{w}}^{i}$ and $\overline{\boldsymbol{z}}^{i}$ are the mean over the n training years of $\boldsymbol{W}$ and $\boldsymbol{Z}$, $\boldsymbol{S}_{ZZ}^{2,i}$ and $\boldsymbol{S}_{WW}^{2,i}$ are the $nmxnm$ covariance matrix of $\boldsymbol{Z}$ and $\boldsymbol{W}$, and $\boldsymbol{S}_{ZW}^{2,i}$ is the $nmxnm$ cross-covariance matrix between $\boldsymbol{Z}$ and $\boldsymbol{W}$. Similarly, the equations (3.17) and (3.18) can be rewritten to solve the FA in the MCA space:

$$\widehat{\boldsymbol{Y}}_i = \overline{\boldsymbol{v}}_b^i + \boldsymbol{\mathcal{L}}^i\left[\widehat{\boldsymbol{w}}_i - \boldsymbol{\mathcal{G}}^i\left(\overline{\boldsymbol{z}}^i - \boldsymbol{z}_0^i\right)\right] \tag{3.31}$$

$$\widehat{\boldsymbol{S}}_i^2 = \left[\left(\boldsymbol{\mathcal{G}}^i\right)^T\left(\boldsymbol{S}^{2,i}\right)^{-1}\boldsymbol{\mathcal{G}}^i + \left(\boldsymbol{\mathcal{C}}^i\right)^{-1}\right]^{-1} \tag{3.32}$$

where $\widehat{\boldsymbol{w}}_i = \widehat{\boldsymbol{m}}_i\boldsymbol{V}_{P,nm}$ is the predictor at the ith target year in the MCA space, $\widehat{\boldsymbol{m}}_i = (\widehat{m}_{i,1} \quad \widehat{m}_{i,2} \quad \cdots \quad \widehat{m}_{i,P})$ is the grid-point climate variable for all M forecast systems at the ith target year and $\boldsymbol{\mathcal{L}}^i = \boldsymbol{\mathcal{C}}^i\left(\boldsymbol{\mathcal{G}}^i\right)^T\left(\boldsymbol{\mathcal{G}}^i\boldsymbol{\mathcal{C}}^i\left(\boldsymbol{\mathcal{G}}^i\right)^T + \boldsymbol{S}^{2,i}\right)^{-1}$ is the gain/weight matrix. The final step is to get the FA predictions in the geographical coordinate system, such as:

$$\widehat{\boldsymbol{Y}}_i^{FAC} = \boldsymbol{U}_{Q,nm}\widehat{\boldsymbol{Y}}_i \tag{3.33}$$

$$\widehat{\boldsymbol{S}}_i^{2,FAC} = \boldsymbol{U}_{Q,nm}\widehat{\boldsymbol{S}}_i^2\left(\boldsymbol{U}_{Q,nm}\right)^T \tag{3.34}$$

where $\widehat{\boldsymbol{Y}}_i^{FAC}$ and $\widehat{\boldsymbol{S}}_i^{2,FAC}$ are the spatial-field predicted mean and covariance matrices of the FA prediction.

3.5. Forecast quality assessment

A forecast quality assessment where the predicted and observed values are compared is an important step in climate prediction. A wide variety of forecast verification procedures exists, but all involve measures of the relationship between a forecast or a set of forecasts and the corresponding observation(s) of the predictand (Wilks, 2006). Due to the high dimensionality of the forecast verification problem, it is very important to take into account multiple verification measures to obtain richer and more robust conclusions about the quality and/or value of the forecast systems (Murphy, 1991; Mason and Stephenson, 2008). In this thesis, several deterministic and probabilistic verification measures are used in the forecast quality assessment. These measures are described below.

The usual Pearson correlation coefficient is used in this thesis to measure the degree of linear correspondence (or association) between the predicted mean (e.g., ensemble mean) and the observed value of the predictand. It can be written as:

$$r = \frac{\sum_{i=1}^{N}\left(\hat{y}_i - \bar{\hat{y}}\right)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}\left(\hat{y}_i - \bar{\hat{y}}\right)^2\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{3.35}$$

where $\hat{y}_i$ and y_i are the predicted and observed predictands and $\bar{\hat{y}}$ and $\bar{y}$ are their respective mean over N the target years. Note that $\hat{y}_i$ and y_i are anomaly values computed in three-year-out cross-validation mode and that their time average in (3.35) are computed for all N target years (i.e., they are close to, but not exactly, zero).

The correlation coefficient measures the quality of deterministic forecasts (i.e., how well the mean of the PDF is predicted), but provides no information about the quality of the forecast uncertainty (i.e., how well the spread of the PDF is predicted). As mentioned in Chapter 2, climate forecasts should be communicated with a proper quantification of the forecast uncertainty in their statements. In this thesis, we use multiple probabilistic verification measures to assess how well the forecast uncertainty is predicted by the forecast systems and their combination. The Brier score (BS) is one of the most commonly used quadratic score to verify probabilistic forecasts. It can be defined as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - o_i)^2 \tag{3.36}$$

where $\hat{p}_i$ is the probability forecast and o_i is the observation, which is set to be one if the event happened and zero if it did not happen, for the ith target year. The BS could be generalized in the form of a skill score where the forecast of a given system is compared to a reference prediction system, which is usually a much simpler forecast such as the climatological frequency of the event. This generalization is called the Brier skill score (BSS), and could be written as:

$$BSS = 1 - \frac{BS}{BS_c}, \qquad with \quad BS_c = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_c - o_i)^2 \tag{3.37}$$

where BS is the Brier score of a given forecast system, BS_c is the Brier score of the climatology and $\hat{p}_c$ is the climatology probability. Positive BSS means the BS of the system is better than the BS of the reference forecast. Two probability events are considered in this thesis: the probability of the predictand of being above the median and the upper quartile.

Two types are probability forecasts are handled in this study: ensemble predictions and sets of predictions defined by a forecast mean and a standard deviation. For those forecast systems that did not have ensemble hindcasts, the normal forecast distribution of each year is sampled with size 10,000 to obtain samples from which to compute the median and the upper quartile of the corresponding climatological distributions. The 10,000 sample size was chosen because it was found to provide robust estimates of the climatological PDF. The robustness was estimated by calculating the BSS 1,000 times for the statistical model and for a given target month and lead time pair. These 1,000 estimations were performed with sample size 11, 51, 100, 1,000 and 10,000. The sample size 10,000 was chosen because it presented the smallest spread in the histogram of the 1,000 estimated values of the BSS. The median and the upper quartile of the climatological distribution are estimated using ensemble members for the predictions of

the dynamical forecast systems and all available years. Separate threshold estimates are obtained for the predictions and the observations to take into account that the predictions have systematic errors in the variability.

Finally, the probability forecasts are estimated using the estimated thresholds (median and upper quartile). For those forecast systems that did not have ensemble hindcasts, the probability forecasts are estimated using the standard PDF for the Gaussian distribution, such as:

$$f(\chi) = \frac{1}{\hat{\sigma}_i\sqrt{2\pi}} exp\left[-\frac{(\chi - \hat{y}_i)^2}{2\hat{\sigma}_i{}^2}\right], \qquad -\infty < \chi < \infty \tag{3.38}$$

where $\hat{y}_i$ and $\hat{\sigma}_i$ are the predicted mean and standard deviation and $\pi = 3.14$ is a mathematical constant. The probability forecast is the area below of the PDF estimated using (3.38) for the two binary events analyzed: χ above the median and the upper quartile. For ensemble hindcasts, the probability forecast is estimated using a frequentist approach. In this case, the probability is given by the number of ensemble members above the threshold divided by the total number of ensemble members of a given forecast system.

The BS can be expressed in terms of the sum of three important forecast verification attributes: the reliability, the resolution and the uncertainty (Wilks, 2006). One way to estimate these three attributes is to stratify the N forecast probabilities into a set of L bins. In this thesis, the forecasts are stratified into 10 bins with equivalent width, ranging from $\hat{p}_i = 0.0$ to $\hat{p}_i = 1.0$ (i.e., $0.0 \leq \hat{p}_i < 0.1$; $0.1 \leq \hat{p}_i < 0.2$, …, $0.9 \leq \hat{p}_i \leq 1.0$). Then, the equation (3.36) can be rewritten as:

$$BS = \frac{1}{N}\sum_{l=1}^{L} N_l(\bar{p}_l - \bar{o}_l) - \frac{1}{N}\sum_{l=1}^{L} N_l(\bar{o}_l - \bar{o}) + \bar{o}(1 - \bar{o}) \tag{3.39}$$

where $L = 10$ is the number of bins, N_l is the number of forecast-observation pairs inside each bin, such as:

$$N = \sum_{l=1}^{L} N_l \tag{3.40}$$

$\bar{p}_l$ and $\bar{o}_l$ are the frequency of predicted and observed events, such as:

$$\bar{p}_l = \frac{1}{N_l}\sum_{i \in N_l} \hat{p}_i \tag{3.41}$$

$$\bar{o}_l = \frac{1}{N_l}\sum_{i \in N_l} o_i \tag{3.42}$$

where o_i is the observation inside each bin. Finally, the frequency of the event being observed in the whole sample (i.e., sample climatology) can be written as:

$$\bar{o} = \frac{1}{N}\sum_{i=1}^{N} o_i = \frac{1}{N}\sum_{l=1}^{L} N_l \bar{o}_l \tag{3.43}$$

The three terms in the equation (3.39) are known as the reliability (BS_{REL}, first term on right hand side of 3.39), resolution (BS_{RES}, second term on right hand side of 3.39) and uncertainty (BS_{UNC}, third term on right hand side of 3.39) components of the BS. BS_{REL} verifies the degree of correspondence between the frequency of events predicted by the system and the frequency of events that actually happened and measures the degree of trustworthiness of the predicted probabilities. On the other hand, BS_{RES} measures the ability of the forecasts to distinguish events that have forecast probabilities different from the climatological frequency given by the equation (3.43). BS_{UNC} is associated with the uncertainty of the observations for a given event being forecast and does not depend on the predictions. For example, an event that always happen and an event that never happens do not have any uncertainty in its observations (i.e., $BS_{UNC} = 0$) as it is fully known the observed outcome. The largest possible uncertainty is achieved for an event that has the climatological frequency equals to 0.5, resulting in $BS_{UNC} = 025$.

Depending on the number of bins used to stratify the forecast probabilities, the sum of the three components of the BS in the equation (3.39) does not equal the BS computed using the equation (3.36). Therefore, two additional components that account for the within-bin variance of the forecasts and the within-bin covariance between forecasts and observations are also needed to make the components of the BS less sensitive to the number of bins (Stephenson et al., 2008). These two extra components are added to the resolution component of the BS to make a generalized resolution term. The generalized resolution can be written as (Stephenson et al., 2008):

$$\begin{aligned} BS_{GRES} = {} & \frac{1}{N}\sum_{l=1}^{L} N_l(\bar{o}_l - \bar{o}) \\ & - \frac{1}{N}\sum_{l=1}^{L} N_l \left[\frac{1}{N_l}\sum_{i \in N_l} (\hat{p}_i - \bar{p}_l)^2 - \frac{2}{N_l}\sum_{i \in N_l} (\hat{p}_i - \bar{p}_l)(o_i - \bar{o}_l) \right] \end{aligned} \tag{3.44}$$

where BS_{GRES} is the generalized resolution component of the BS. In this thesis, the skill scores of the reliability and generalized resolution are used with the climatology as reference. They are computed as follows (Doblas-Reyes et al. 2005):

$$BSS_{REL} = 1 - \frac{BS_{REL}}{BS_{UNC}} \tag{3.45}$$

$$BSS_{GRES} = \frac{BS_{GRES}}{BS_{UNC}} \tag{3.46}$$

Another probabilistic verification score used in this thesis is the continuous ranked probability score ($CRPS$; Wilks, 2006). It can be generically defined as:

$$CRPS = \int_{-\infty}^{\infty} [F^*(\chi) - F_0(\chi)]^2 \, d\chi \tag{3.47}$$

where

$$F_0(\chi) = \begin{cases} 0, & \chi < y_i \\ 1, & \chi \geq y_i \end{cases} \tag{3.48}$$

is a Heaviside step function that jumps from 0 to 1 at the point where the predicted variable equals the observation and $F^*(\chi)$ is the predicted cumulative distribution function (CDF). An illustrative example of the $CRPS$ is given in Figure 3.1. The $CRPS$ has two important advantages over the BS (Jolliffe and Stephenson, 2012): it is defined on a continuous scale and does not require reduction to discrete probabilities of binary events, and it can be interpreted as an integral over all possible BS values. On the other hand, the BS allows to verify how a forecast system predict different kind of probability events, such as, more extreme events. Therefore, they are complementary.

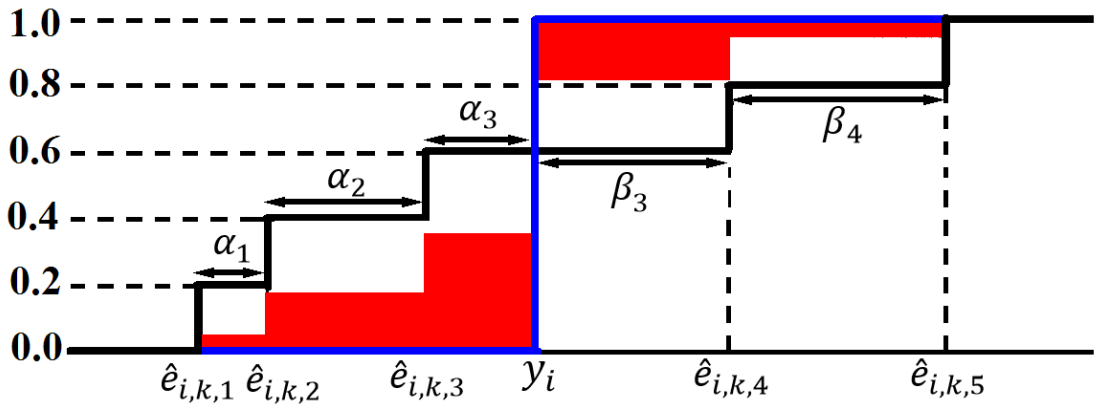

Figure 3.1: Cumulative distribution for the kth forecast system of five members $\{\hat{e}_{i,k,1}, \cdots \hat{e}_{i,k,5}\}$ *(thick solid black line) and the Heaviside step function that jumps from 0 to 1 at the observation* y_i *(thick solid blue line). The CRPS at the ith target year is represented by the red area. Source: Adapted from Hersbach (2000).*

The $CRPS$ is estimated differently for the two types of probability forecasts handled in this thesis. When ensemble predictions are considered (Figure 3.1), the $CRPS$ is estimated assigning equal weight to each ensemble member, such as:

$$\hat{p}_{i,k,t} \equiv \frac{t}{N_t}, \qquad for \quad \hat{e}_{i,k,t} < \hat{e}_{i,k,t+1} < \cdots \hat{e}_{i,k,N_t} \tag{3.49}$$

where $\hat{p}_{i,k,t}$ is a piecewise constant function for the tth ensemble member, N_t is the total number of ensemble member of the kth forecast system at the ith target year. Note that $\hat{p}_{i,k,0} = -\infty$ and $\hat{p}_{i,k,N_t+1} = \infty$ are introduced for convenience. For the hypothetical forecast system illustrated in Figure 3.1, $\hat{p}_{i,k,1} = \cdots = \hat{p}_{i,k,5} = \frac{1}{5}$.

Therefore, the equation (3.47) can be rewritten to deal with ensemble forecasts, as follows (Hersbach, 2000):

$$c_{i,k,t} = \int_{t}^{t+1} \left[\hat{p}_{i,k,t} - F_0(y_i)\right]^2 d\chi \Longrightarrow CRPS = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N_t} c_{i,k,t} \tag{3.50}$$

where

$$c_{i,k,t} = \alpha_t \hat{p}_{i,k,t}{}^2 + \beta_t \left(1 - \hat{p}_{i,k,t}\right)^2 \tag{3.51}$$

If the observation y_i falls between the smallest and highest ensemble member, then α_t and β_t can be estimated as follows:

$0 < t < N_t$	α_t	β_t
$y_i > \hat{e}_{i,k,t+1}$	$\hat{e}_{i,k,t+1} - \hat{e}_{i,k,t}$	0
$\hat{e}_{i,k,t+1} > y_i > \hat{e}_{i,k,t}$	$y_i - \hat{e}_{i,k,t}$	$\hat{e}_{i,k,t+1} - y_i$
$y_i < \hat{e}_{i,k,t}$	0	$\hat{e}_{i,k,t+1} - \hat{e}_{i,k,t}$

(3.52)

Otherwise, when the observation y_i falls outside the range of ensembles, then α_t and β_t can be estimated as follows:

Outlier	α_t	β_t
$y_i < \hat{e}_{i,k,1}$	0	$\hat{e}_{i,k,t} - y_i$
$y_i > \hat{e}_{i,k,N_t}$	$y_i - \hat{e}_{i,k,t}$	0

(3.53)

Note that in these cases, $t = 0$ $\left(\Delta\chi = \hat{e}_{i,k,1} - y_i\right)$ and $t = N_t\left(\Delta\chi = y_i - \hat{e}_{i,k,N_t}\right)$ concern the intervals $(-\infty, \hat{e}_{i,k,1}]$ and $\left[\hat{e}_{i,k,N_t}, \infty\right)$, respectively, for which $\hat{p}_{i,k,0} = 0$ and $\hat{p}_{i,k,N_t} = 1$.

In the case of the statistical model and the combinations, where sets of predictions defined by a forecast mean and standard deviation are considered, the $CRPS$ is estimated as follows (Gneiting et al., 2005):

$$CRPS = \frac{1}{N}\sum_{i=1}^{N} \hat{\sigma}_i \left\{ \frac{y_i - \hat{y}_i}{\hat{\sigma}_i} \left[2F\left(\frac{y_i - \hat{y}_i}{\hat{\sigma}_i} \right) - 1 \right] + 2f\left(\frac{y_i - \hat{y}_i}{\hat{\sigma}_i} \right) - \frac{1}{\sqrt{\pi}} \right\} \tag{3.54}$$

where

$$F(\chi) = \frac{1}{\sqrt{2\pi}} \int exp\left[-\frac{\chi^2}{2} \right] d\chi \tag{3.55}$$

where $F(\chi)$ and $f(\chi)$ denote the CDF and PDF, respectively, of the normal distribution with zero 0 and variance 1 evaluated at the normalized prediction error, $\frac{y_i - \hat{y}_i}{\hat{\sigma}_i}$.

The $CRPS$ can be computed in terms of skill score, such as:

$$CRPSS = 1 - \frac{CRPS}{CRPS_{CLIM}} \tag{3.56}$$

where $CRPS_{CLIM}$ is the $CRPS$ of the climatological distribution computed using the equation (3.50) and considering all but the target year with equal probability, $\hat{p}_{i,1} = \cdots = \hat{p}_{i,N-1} = \frac{1}{N-1}$.

Another probabilistic score used in this thesis is the ignorance score (Ign), which is defined as the negative logarithm of the predicted probability density corresponding to the event that actually occurred (Wilks, 2006). It can be written as:

$$Ign = \frac{1}{N}\sum_{i=1}^{N} -ln\left[\hat{f}(y_i)\right] \tag{3.57}$$

where $\hat{f}(y_i)$ is the predicted probability density of the observed predictand y_i. The Ign measures the information deficit (or ignorance) of a forecast probability when the observation is available (Roulston and Smith, 2002). That is, what information is missed for the probability forecast equals the observation. For a perfect forecast, 100% forecast probability has to be issued for the event that actually occurred. In this case, the ignorance score would be 0. Alternatively, Ign goes to infinity when 0% forecast probability is assigned for the event that actually occurred. That means that the Ign punishes very hard bad probabilistic forecasts (Gneiting et al., 2005). Ensemble predictions, especially for those dynamical forecast systems that have a small number of ensemble members, usually get low evaluation in terms of Ign.

As previously, the predicted PDF is estimated in two different ways. When ensemble prediction is available $\hat{f}(\chi)$ is estimated using a kernel density estimator, which sums of the bumps placed at each ensemble member for a given forecast system. For simplicity, each bump will be assumed Guassian. For the statistical model and the combinations, the predicted PDF is assumed Guassian and the equation (3.38) is used to solve the Ign. After some arithmetic, the equation (3.57) can be rewritten as (Gneiting et al., 2005):

$$Ign = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{2}ln\left(2\pi\hat{\sigma}_i^{\,2}\right) - \frac{(y_i - \hat{y}_i)^2}{2\hat{\sigma}_i^{\,2}}\right] \tag{3.58}$$

Rewriting the Ign in terms of skill score so that:

$$IgnSS = 1 - \frac{Ign}{Ign_{CLIM}} \tag{3.60}$$

where Ign_{CLIM} is the ignorance score of the climatological distribution, estimated considering all but the target years as bumps in the kernel estimation, and Ign is the ignorance score of the system being evaluated.

A particular verification dataset is just one of many possible samples from a population and therefore verification measures need to be shown together with an indication of the sampling uncertainty (Jolliffe and Stephenson, 2012). The sampling uncertainty in the verification measures is quantified using 95% confidence intervals (Nicholls, 2001; Mason, 2008b; Jolliffe and Stephenson, 2012). However, for grid-point verification measures displayed as a map or two-dimensional field where the use of confidence intervals would result in a very complex map, p-values are used to quantify the sampling uncertainty (Nicholls, 2001). Both the confidence intervals and the p-values are estimated using a non-parametric bootstrap method (Mason, 2008b; Jolliffe and Stephenson, 2012). In this procedure, the forecast-observation pairs are randomly resampled with replacement, keeping the forecast and observation pairs together (Mason, 2008b). The bootstrap size is chosen to be 1,000. From these 1,000 resamples, the 2.5% and 97.5% quantiles, which represent the lower and upper confidence interval limits, respectively, are estimated. On the other hand, the null hypothesis used to estimate the p-values is that the verification measure is zero, while the alternative hypothesis is that it is larger than zero (i.e. one-tailed test). When the difference between the forecast quality of two forecast systems are considered, the alternative hypothesis is that the verification measure difference is different from zero (i.e., two-tailed test).

4. Prediction of tropical SST

4.1. Introduction

Due to the chaotic nature of the climate system and the inadequacy of current dynamical forecast systems, quantifying uncertainty plays an important role in climate forecasting (Palmer, 2000). Dealing with uncertainty will help decision makers making better decisions on whether or not to take any action given a probability forecast for an event. The unavoidable uncertain character of weather and climate prediction forces climate forecasts to be formulated in a probabilistic way, as has been recognized for more than a century (Murphy and Winkler, 1984). In addition, the probabilistic formulation requires an appropriate assessment of how reliable (i.e. whether the forecast uncertainty is accurate) the forecasts are (Slingo and Palmer, 2011).

The predictability of the near-surface temperature and precipitation patterns, which plays an important role on human activities, is to a certain degree linked to our ability to predict the boundary conditions of the climate system such as the SST, especially in the tropics (Shukla, 1998; Goddard et al., 2001). As shown in Chapter 2, ENSO is the most important source of predictability at seasonal timescale; therefore, the assessment of skill of ENSO SST predictions is a fundamental requirement for any seasonal forecasting system (Stockdale et al., 2011). Other tropical ocean basins such as the tropical SST over the Atlantic and Indian Oceans also have a major impact on the climate variability of the surrounding regions (Goddard et al., 2001). For instance, the SST anomalies over the tropical Atlantic region directly influence the position of the ITCZ, which plays a role on the precipitation patterns over northern northeastern Brazil and western Africa, while the western Indian Ocean SST anomalies have impacts on the climate of eastern parts of the African continent. Another interesting feature is that the SST variability of the Atlantic and Indian basins is somehow linked to that of the tropical Pacific (Goddard et al., 2001). Therefore, an important tool used for operational seasonal predictions are the ocean climate indices that can be linked to major patterns of climate variability (Doblas-Reyes et al. 2013a).

This chapter addresses several innovative aspects of climate forecasting. Firstly, it compares three different operational dynamical forecast systems: S4, CFSv2 and MF3. These are some of the dynamical seasonal forecast systems available to the users of this type of climate information. A simple statistical model based on lagged regression (Coelho et al., 2004) is also used as an additional model in the combination procedure. Secondly, the study uses and compares several methods to combine the single forecast systems in different ways: the multiple linear regression methods described in Doblas-Reyes et al. (2005) and the Bayesian method described in Stephenson et al. (2005). The SMM, where the systems are put together with equal weighting, is used as a benchmark. The aim is to assess how the Bayesian method compares with the multiple linear regression methods and a simple multimodel. However, this study goes a bit farther than those two papers, and several others that were recently published (Hagedorn et al. 2005; Palmer et al. 2004; Tippett and Barnston 2008; Kug et al. 2007; Kug et al. 2008). In this

paper, the impact of those combination methods on a series of operational dynamical forecast systems, which is an aspect of the problem not dealt with in the past, was investigated. In particular, this implied considering the differences in how the systems are developed in a real-time basis, and how the combination affects predictions that are carried out regularly, with one start date per month. Finally, a comprehensive quality assessment of the climate forecasts both from a deterministic and probabilistic point of view is performed considering all possible start dates and lead time up to seven months, which is the limit of the forecast time allowed by both S4 and MF3. The forecast quality assessment of both the combinations and the single forecast systems is carried out for SST averaged over three different tropical regions: the Niño3.4, the SNA, and the WTI indices.

4.2. Forecast quality assessment

4.2.1. Niño3.4 index

Figure 4.1 shows monthly forecast anomalies of the Niño3.4 index for the four single forecast systems and the FAS for the period between 1982 and 2010. This illustration is for the target month of January and lead time two months. This means that the statistical model used the previous month of October of the previous year as the predictor, S4 and MF3 forecasts were started on the first of November while CFSv2 has its ensemble members started between the second week of October and the first week of November. The 95% prediction interval for each forecast system, given by the predicted mean anomaly plus or minus 1.96 times the predicted standard deviation, and the mean climatology forecast are also displayed. For the dynamical forecast systems the predicted standard deviation is the standard deviation of all available ensembles. All forecast systems have a high linear correspondence with the observed anomalies. However, the FAS has higher correlation than all single forecast systems. Besides, most of the observations in the forecast systems, except for the MF3, fall inside the 95% prediction interval meaning that these forecast systems are reliable. This is shown quantitatively by the large values of the reliability component (BSSrel) of the BSS of each forecast system that is displayed in the top left corner of each panel of Figure 4.1. Note that even though many of the MF3 forecasts fall outside the prediction interval, it has a good BSSrel. The FAS has the highest BSS of all forecast systems.

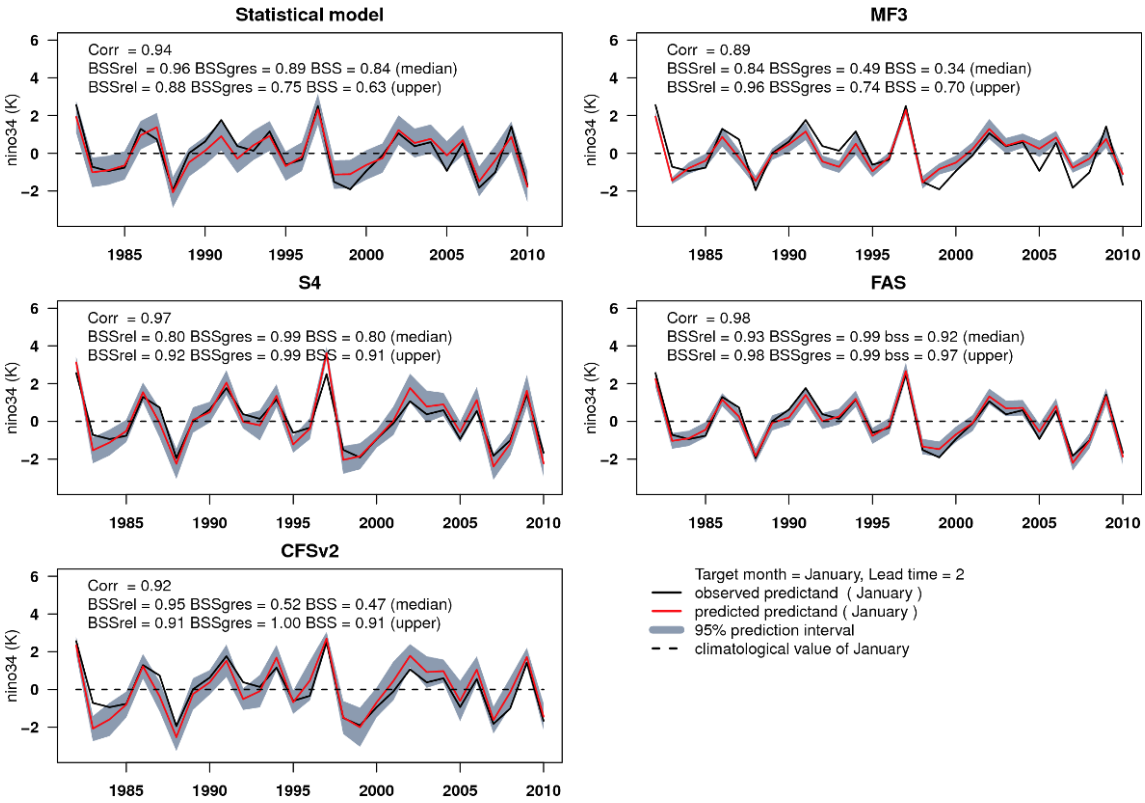

Figure 4.1: Monthly forecast anomalies of Niño3.4 index for the statistical model, S4, CFSv2, MF3 and FAS. Forecasts are for the target month of January with lead time two. Observed values (black solid line), predicted values (red solid line), 95% predicted interval (grey area) and the climatology value of January (black dashed line). Several scores are displayed in each panel: the correlation coefficient, and the Brier skill score and its reliability and resolution components for dichotomous events of SST anomalies exceeding the median and the upper quartile.

As operational systems have to provide a prediction starting at least once a month all year round, this analysis was extended to all months of the year and for the seven lead times available for S4 and MF3. The results for both deterministic and probabilistic scores are summarized in Figures 4.2 and 4.3. Figure 4.2 shows the correlation coefficient of the Niño3.4 SST index mean prediction as a function of both target month and lead time for all forecast systems and combinations for the period between 1982 and 2010. The statistical model has the highest values of correlation for predictions produced during the boreal winter, when ENSO persistence is the strongest, followed by a period of decreasing skill for predictions produced during the boreal spring. This decrease in skill during the boreal spring is known as the spring barrier (Balmaseda et al. 1995; Goddard et al. 2001; Mason and Mimmack 2002; Stockdale et al. 2011). The lowest values of correlation were observed during the boreal summer for longer leads coinciding with the period of the year when ENSO typically changes from one phase to another.

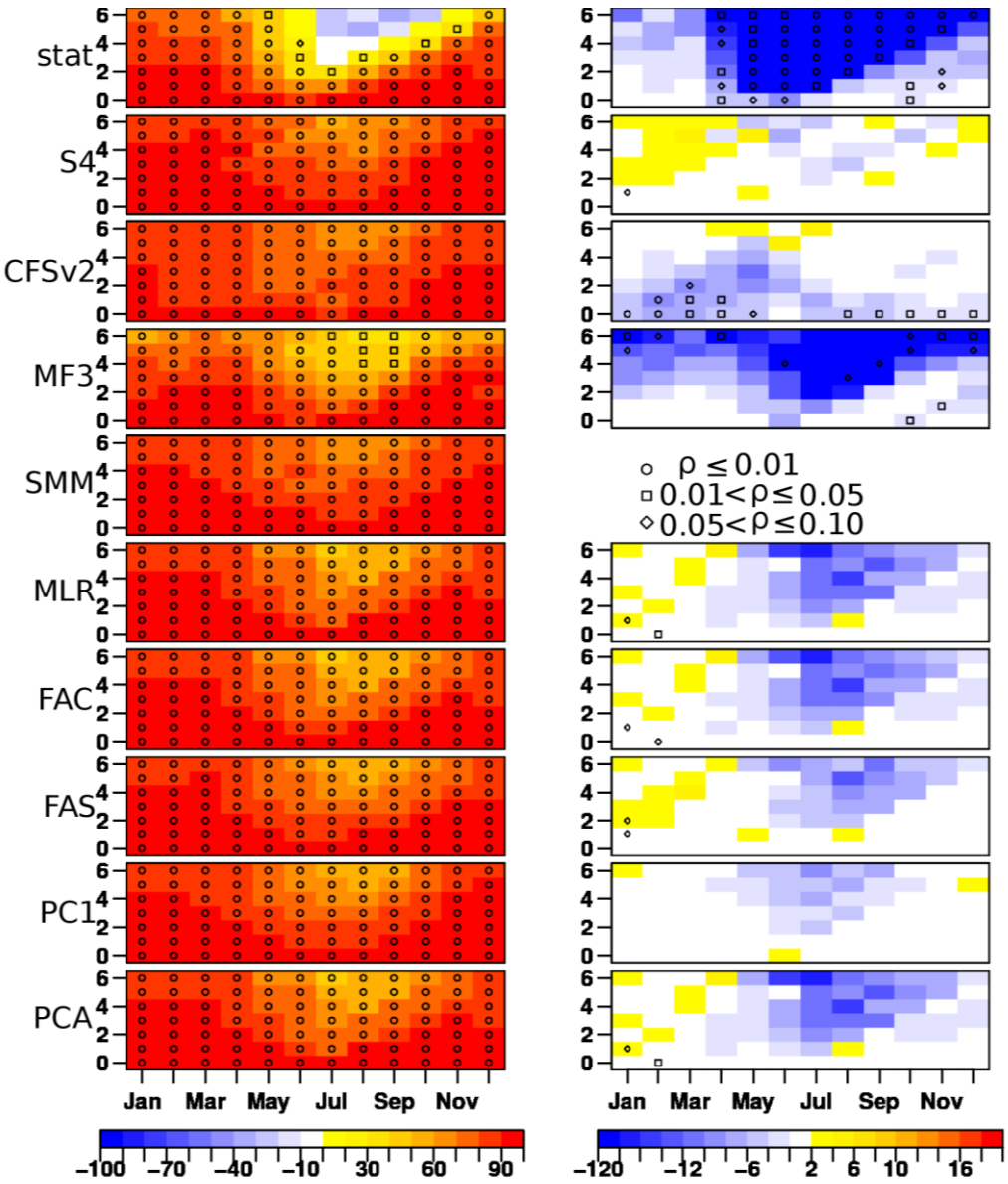

Figure 4.2: (Left column) Correlation between the ensemble-mean predicted and observed Niño3.4 index as a function of target month (horizontal axis) and lead time (vertical axis) for the different forecast systems. (Right column) Correlation difference between each forecast system and the SMM. The predictions have been formulated over the period 1982 to 2010. The forecast systems used are, from top to bottom the statistical model, S4, CFSv2, MF3, SMM, MLR, FAC, FAS, PC1 and PCA-regression. HadISSTv1.1 data are used to estimate the coefficients in the statistical model and for the forecast quality assessment. Circles are for p-values smaller than or equal 0.01, squares for p-values between 0.05 and 0.01, and diamonds for p-values between 0.10 and 0.05.

A similar pattern is found for the S4 predictions, except that the correlation is higher, and the decrease of skill for predictions started during the boreal spring is much less important than in the statistical predictions. The superior performance of the ECMWF seasonal forecast system 3 (S3) over the older ECMWF forecast systems and persistence when predicting the Niño3.4 index has been shown previously (Stockdale et al., 2011). S4 and

S3 have similar skill in terms of anomaly correlation when predicting the Niño3.4 index (Molteni et al., 2011). The CFSv2 predictions also show a less marked decrease in correlation across the spring barrier than the statistical model, but on average its skill is slightly lower than in S4. Kim et al. (2012) also found that S4 has higher correlation than CFSv2 when predicting the Niño3.4 index in the boreal winter with lead time one month; however, they did not apply the CFSv2 bias correction suggested by Kumar et al. (2012). Here we show that S4 outperforms CFSv2 even after applying the bias correction suggested by Kumar et al. (2012). On the other hand, MF3 predictions are less skilful than the other dynamical forecast systems and also have a decrease in correlation during the boreal spring although its correlation does not turn into negative correlation as in the statistical model. Similar results were found for the NCEP CFS version 1 (CFSv1) and a persistence model in previous studies (Saha et al., 2006; Sooraj et al., 2012). It is worth noting that ENSO skill may have a decadal dependence, i.e. skill may depend on the period of verification (Balmaseda et al., 1995). That is not addressed here. Moreover, persistence forecasts, though outperformed by more sophisticated statistical models of ENSO, are a tough standard to beat mainly when predicting short lead times (Goddard et al., 2001; Mason and Mimmack, 2002). In any case, none of the three dynamical forecast systems as well as none of the combinations, show any negative correlation as the statistical model does in boreal summer.

The SMM, which is used as the reference standard in the comparison with all the other forecast systems, has higher correlation than the statistical model, CFSv2 and MF3 more often than not (right panel of Figure 4.2). On the other hand, the SMM has higher correlation than S4 only at longer leads in the boreal summer and fall. As discussed in previous studies (Hagedorn et al., 2005) the SMM could be outperformed by the best single forecast system on some of the aspects of the prediction (here the Niño3.4 index in the boreal winter). On the other hand, as it will be shown in the following sections of this chapter, the SMM has an overall better performance than the four single forecast systems when all aspects of the prediction (i.e. the three analyzed regions, all target month and lead time pairs) are taken into account.

All combinations show a similar skill pattern. They outperform the SMM only in a few target month and lead time pairs especially during the boreal winter. On the other hand, they have lower correlation than the SMM in the other months of the year, especially for leads longer than two months. The forecast quality of the SMM is usually difficult to improve using multiple linear regression because of the small number of single forecast systems and short time series used to estimate the regression coefficients (Doblas-Reyes et al., 2005). This could also help explaining the similarities between the MLR and PCA predictions. S4 has the best overall correlation for the Niño3.4 predictions, that is, it has higher correlation than all the other single forecast systems and combinations more often than not.

Because of the inherent uncertainty involved in climate forecasting (Mason and Mimmack, 2002) the quality of the probabilistic forecasts were also assessed and will be described below. The BSS with respect to climatology for the SST anomalies exceeding the median for the Niño3.4 index is shown in Figure 4.3. The patterns of skill are similar to those of the correlation coefficient, except for the smaller magnitude of the values. Wang et al. (2009) also found that probabilistic forecast skill scores display similar patterns as those of deterministic scores. Positive BSS for most of the target month and lead time pairs in all single forecast systems and their combinations show that predictions are more skillful than the climatology. As for the mean predictions, the probabilistic predictions for this index also show the lowest skill at the target period of the boreal summer, especially for longer leads (i.e. for predictions with start dates in boreal spring). This agrees with previous studies (Tippett and Barnston, 2008).

The differences between the BSS of the single forecast systems and the combinations with the SMM have all a pattern similar to that of the correlation shown in Figure 4.2. One difference is that the SMM beats all single forecast systems, including S4, more often than not. On the other hand, the other combinations are more competitive when assessing their probabilistic skill in comparison with their deterministic counterpart although the SMM performs better than all of them more often than not. The only exception to this is given by the FAS predictions. The only season where the unequal combinations beat the SMM more often than not is the boreal winter. This is achieved by improved resolution skill score, a highly desirable feature that shows that unequal weighting can also improve the accuracy of the predictions and not just the reliability. This result provides evidence that unequal combination can indeed improve predictions over that of the SMM even with the limited sample size typical of seasonal forecasting.

The BSS for the event defined as the anomalies of the Niño3.4 SST index exceeding the upper quartile were also analyzed (not shown). The BSS has similar patterns to those of the SST anomalies exceeding the median shown in Figure 4.3. One difference is that the SMM is more difficult to beat when predicting the SST anomalies exceeding the upper quartile.

4.2.2. Subtropical North Atlantic index

The correlation coefficient of the predictions of the SNA SST index for all forecast systems and combinations show positive values in all target month and lead time pairs (Figure 4.4). This is also observed in the Niño3.4 index predictions, except for the statistical model where negative correlations are observed in the boreal summer and beginning of fall at leads four, five and six. This confirms that there is considerable SST memory in these two ocean regions for the lead times considered here. For all forecast systems and combinations the correlation coefficient is higher in the Niño3.4 index than in the SNA index more often than not. All these findings agree with previous studies, although they used slightly different areas to represent the tropical northern Atlantic SST region (Stockdale et al., 2011; Sooraj et al., 2012;).

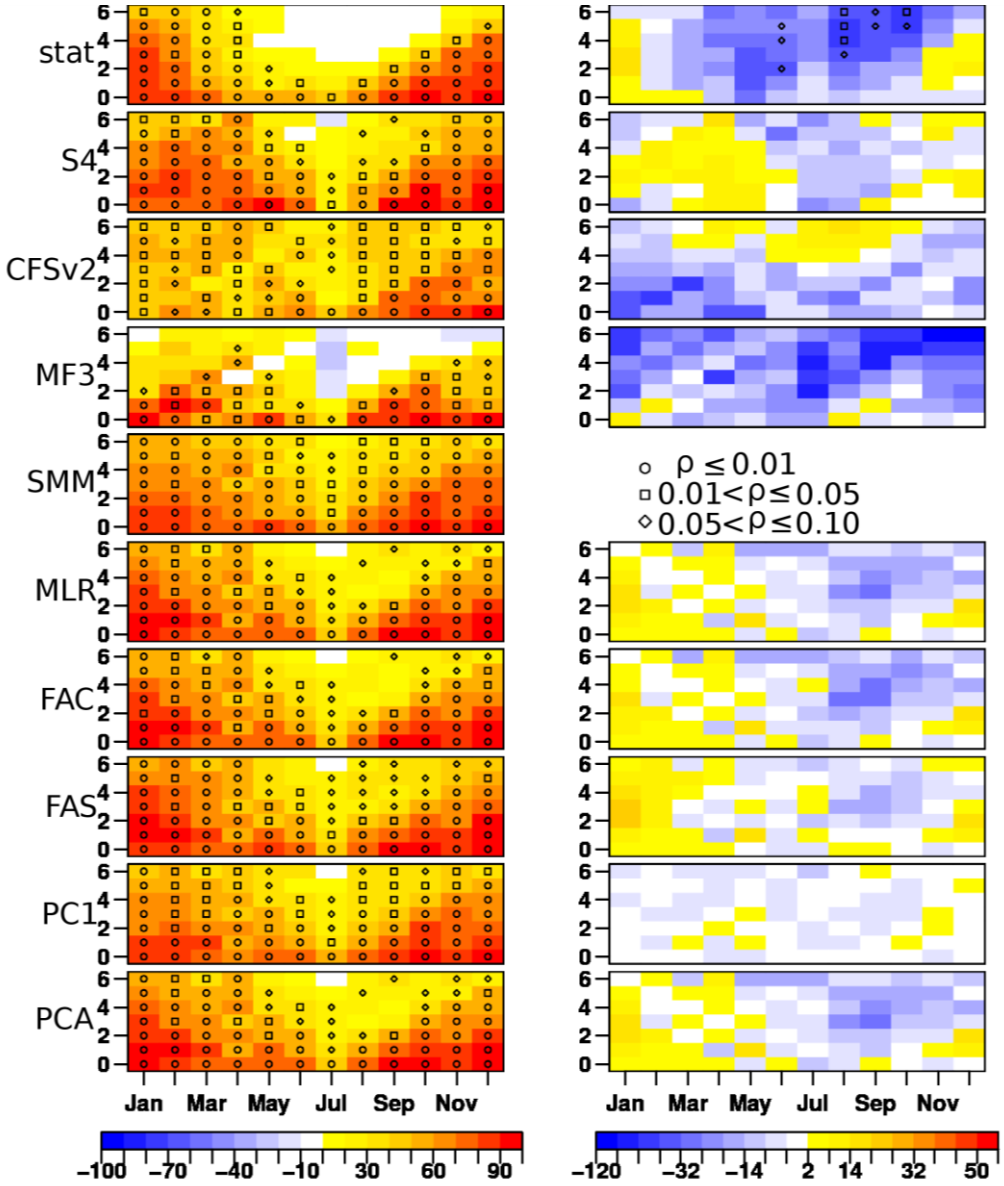

Figure 4.3: As Figure 4.2, but for the BSS of the Niño3.4 SST index anomalies exceeding the median.

The skill of the SNA index varies seasonally. All forecast systems reach a maximum peak in correlation in December. After the peak the correlation starts decreasing, reaching relatively lower values of correlation during boreal spring. The SMM has higher correlation than all forecast systems and combinations during all seasons more often than not, except for the S4 and CFSv2 in the boreal summer and fall (right panel of Figure 4.4). This shows how difficult it is to improve the SMM ensemble-mean predictions in all cases using more sophisticated combination methods that assign unequal weights to each forecast system (Doblas-Reyes et al., 2005; DelSole et al., 2011).

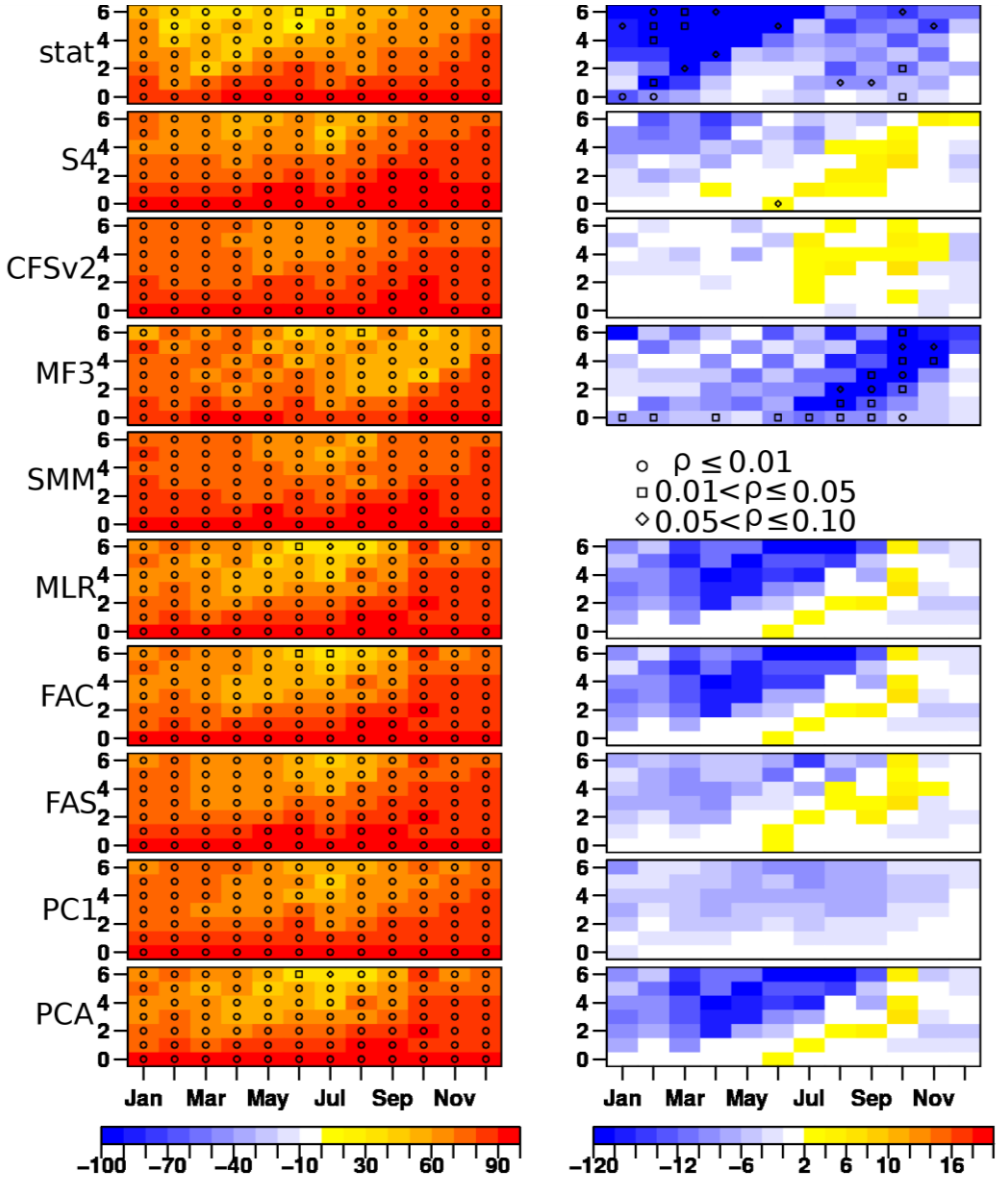

Figure 4.4: Same as Figure 4.2, but for the correlation coefficient of the SNA SST index anomalies.

The BSS of the SNA SST index anomalies exceeding the median have similar patterns as the correlation counterpart, except for the lower magnitude (Figure 4.5). The SMM has higher BSS than the statistical model, S4 and MF3 more often than not. CFSv2 beats the SMM in terms of BSS more frequently than not, but S4 beats the SMM only during the boreal summer and fall in some leads. It is important to note that S4 shows noticeable improvements in skill when compared to S3 and persistence over the tropical Atlantic region (Molteni et al., 2011). The SMM outperforms all combinations methods more often than not; however, it is observed that the unequal combinations do a good job during the target months between August and November. The decomposition of the BSS shows that the resolution skill score term explain most of the pattern of the BSS in all systems, that is, whenever the SMM has higher (smaller) BSS than a single forecast system or combination it also performs better (worse) in terms of BSSres. All forecast systems and

combinations have similar BSSrel, except for the SMM that performs better in a few target month and lead time pairs.

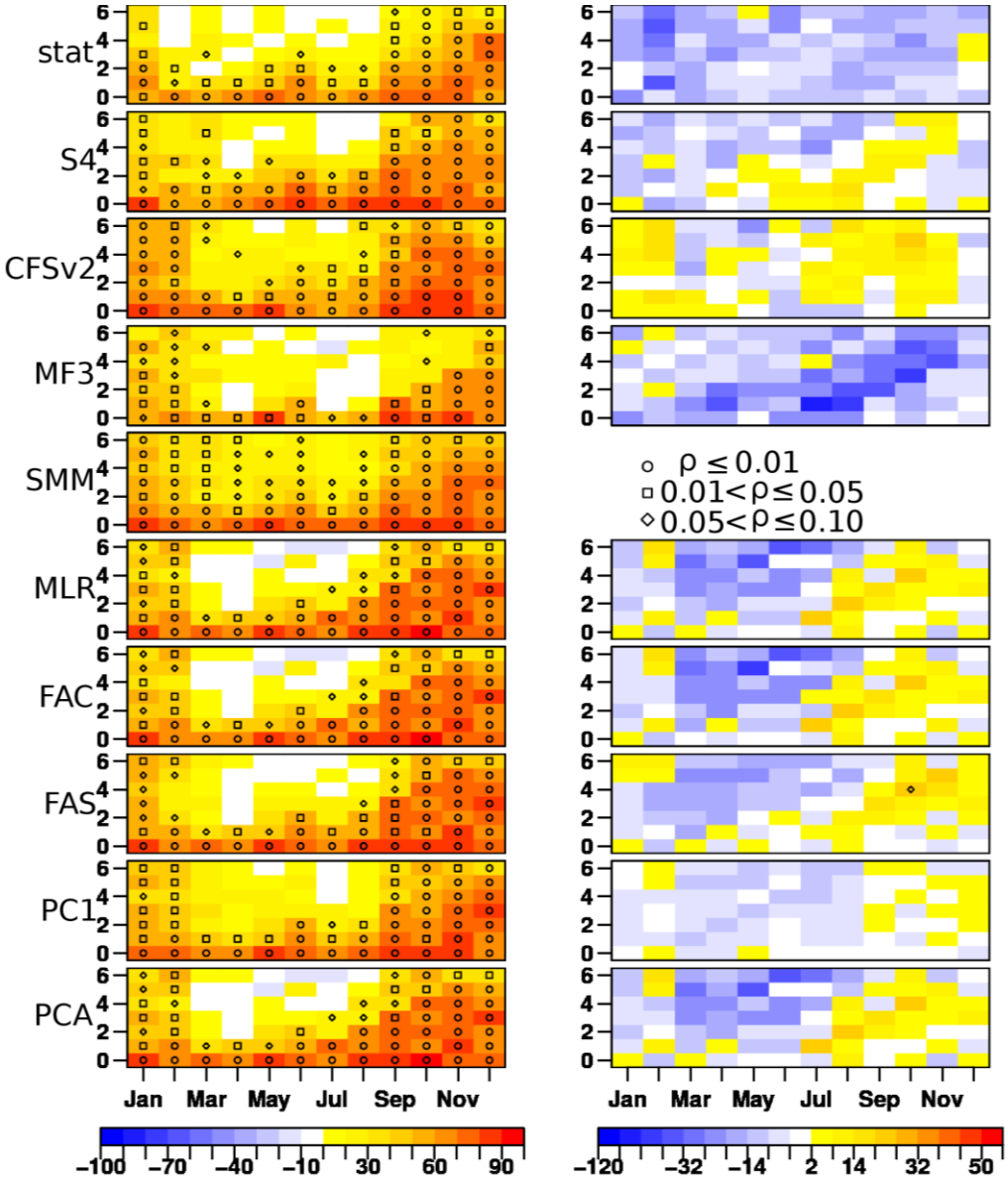

Figure 4.5: As Figure 4.2, but for the BSS of the SNA SST index anomalies exceeding the median.

The BSS of the SNA SST index anomalies exceeding the upper quartile shows similar patterns as the ones in Figure 4.5, except that they are smaller in magnitude (not shown). As for the Niño3.4 index, predictions of the SNA SST index anomalies exceeding the upper quartile are less skillful than when predicting the event of exceeding the median. In addition, in agreement with the results discussed above, the SMM has higher BSS than all forecast system more often than not. For the Niño3.4 and SNA indices it is more difficult to improve SMM forecasts in terms of BSS for more extreme events, such as the ones above the upper quartile than for events above the median.

4.2.3. Western Tropical Indian index

All forecast systems show positive correlation for the WTI index predictions in almost all target month and lead time pairs (Figure 4.6). This shows both that there is considerable SST memory in the region and that the three dynamical forecast systems analyzed here are able to reproduce well the inter-annual SST variability in the Western Indian Ocean. The predictability of the WTI index also varies seasonally, but unlike the two indices described above the skill of the statistical and the three dynamical forecast systems vary differently. The three dynamical forecast systems have the highest correlation during the target months between November and May, and a significant drop in skill in the target months of the boreal summer, especially for longer lead times. This rapid decrease in correlation was observed previously in S3 (Stockdale et al., 2011), CFSv1 (Sooraj et al., 2012), the Climate Prediction and its Application to Society (CliPAS) and the DEMETER multimodel (Wang et al., 2009). On the other hand, the statistical model has two peaks in correlation, one in the boreal spring and another one in the boreal summer. Wang et al. (2009) showed that while the SST predictions in the WTI and East Tropical Indian (ETI; 90°-110°E, 10°S-Equator) have some useful skill both in dynamical and statistical forecast systems and their combinations, the skill for the Indian Ocean Dipole SST index (SST at ETI minus SST at WTI; Saji et al., 1999), which has influence in the surround continental regions, is reduced.

The statistical model has higher correlation than the SMM at the first two lead times in all target months and during the boreal summer also at longer leads when the three dynamical forecast systems have relatively lower skill. CFSv2 is the only dynamical forecast system that outperforms the SMM more often than not. On the other hand, S4 and MF3 have systematically lower correlation than the SMM. All combinations, except for the PC1, outperform the SMM more often than not and this coincides with the target month and lead time pairs where the statistical model performs well. This shows that in some situations the combination methods that assign unequal weights can in fact lead to improvement in skill over that of the SMM. In contrast to the Niño3.4 and SNA analyses, the inclusion of the statistical model information in the WTI index adds skill to the combinations.

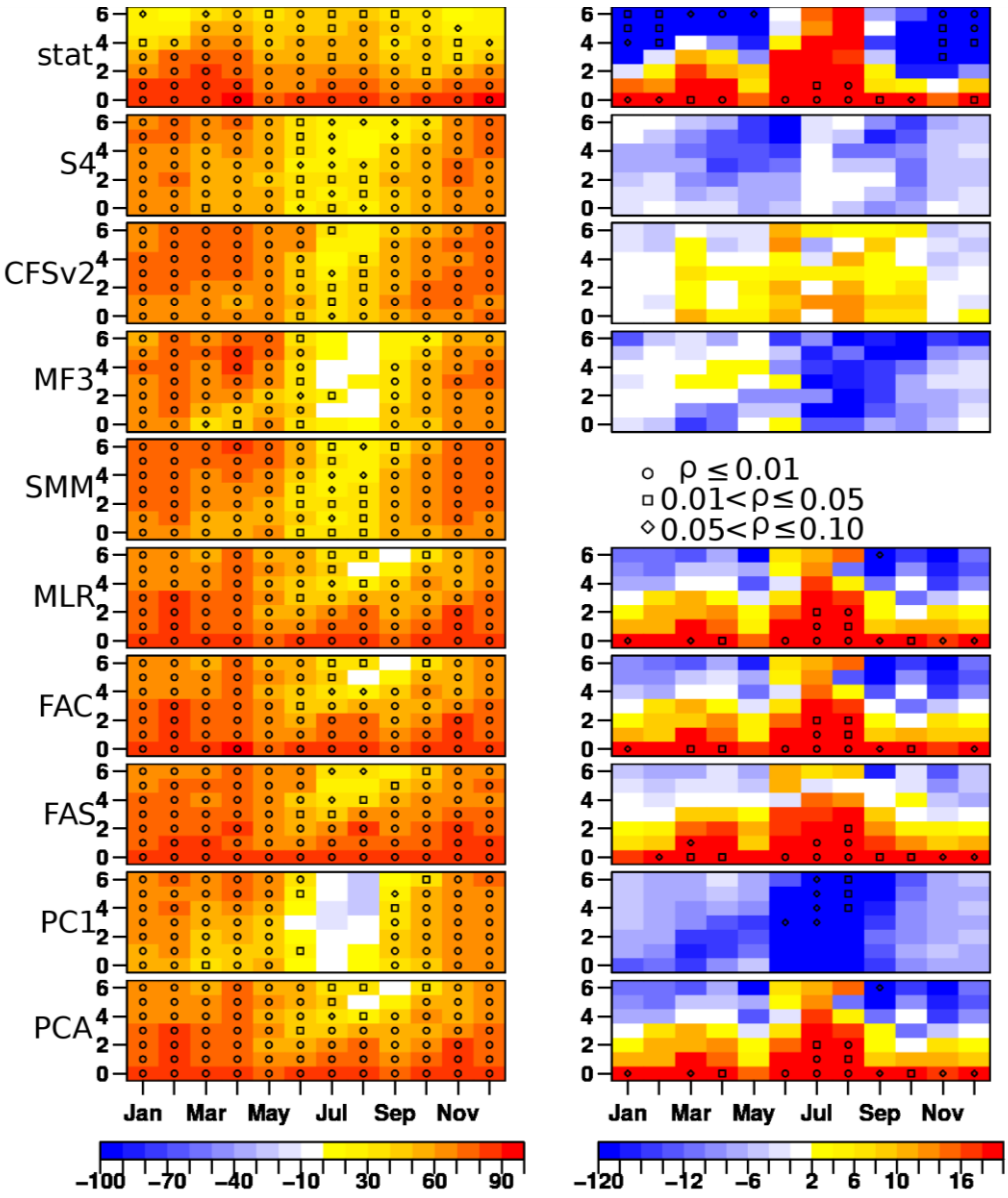

Figure 4.6: As Figure 4.2, but for the correlation coefficient of the WTI SST index anomalies.

The BSS of the WTI SST index anomalies exceeding the median are shown in Figure 4.7. Except for the statistical model during the boreal summer, when the dynamical forecast systems perform worse than climatology, the SMM outperforms all forecast systems more often than not. The statistical model has higher BSS than the SMM during the target months of July and August at all lead times and also during the first two leads in the first six target months of the year. For the other target month and lead time pairs the SMM has higher BSS than the statistical model more often than not. The SMM has systematically higher BSS than all dynamical forecast systems in all seasons of the year and lead times. Among all forecast systems and combinations the FAS is the only one that has higher BSS than the SMM more often than not. On the other hand, the PC1 is the only combination that has systematically lower BSS than the SMM while the other combinations have higher BSS than the SMM at the target month and lead time pairs

when the statistical model performs well. The decomposition of the BSS shows that the combination methods that assign unequal weights have higher reliability skill score than the SMM more often than not, but they only perform better than the SMM in terms of resolution skill score during the boreal summer (not shown).

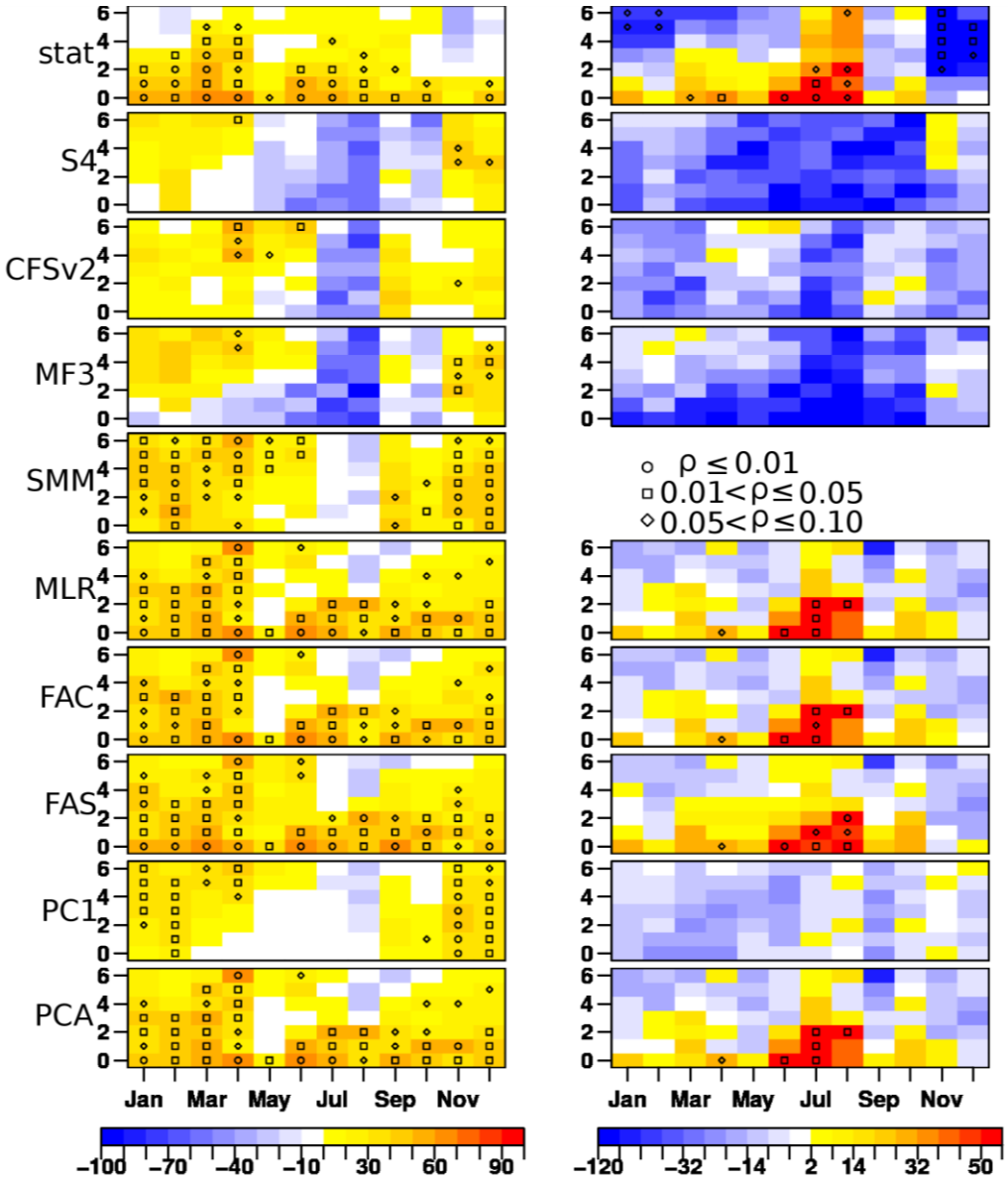

Figure 4.7: As Figure 2, but for the BSS of the WTI SST index anomalies exceeding the median.

The BSS of the WTI SST index anomalies exceeding the upper quartile shows similar results as those of Figure 4.7. However, all dynamical forecast systems perform worse than the climatology more often than when predicting the same index exceeding the median. The boreal summer is the most difficult season to improve the climatological probability forecasts and the statistical model is the only single forecast system that has skill. All combinations also have skilful probabilistic predictions when predicting the

WTI index anomalies exceeding the upper quartile. Besides, the WTI is the only index when it is easier to outperform the SMM when predicting events above the upper quartile.

4.2.4. Discussion

The previous results give a detailed account of the different performance of the single forecast systems and the impact of the combination methods. It was seen that there is no single forecast system that provides the best results for all cases. In fact, while one system is better for Niño3.4 (S4), a different one is the best overall for the WTI (statistical). Surprisingly, simple empirical models can still provide useful predictive information, even when compared to the recently developed state-of-the-art dynamical forecast systems. As it is impossible to choose a single system to provide climate information to the users, an approach to integrate the different sources into a single prediction is necessary. These combination methods have different properties for specific target month and lead time pairs, not only because they combine a different set of single systems, but also because they calibrate the probabilistic predictions differently, as it has been found in the analysis of the reliability.

A more integrated view of the advantages of the set of combination methods considered is required. The scatterplots of the correlation (Figure 4.8) and the BSS (Figure 4.9) summarize the results described above. Predictions for all indices, target months, lead times and events of the probabilistic forecasts have been included to obtain a general picture of the performance. Each symbol in the scatterplots represents one of the three analyzed regions: WTI (circle), Niño3.4 (triangle) and SNA (cross).

The SMM has higher correlation than all single forecast systems and combinations, except for the FAS, more often than not (not shown). This is seen in Figure 4.8 by the number of symbols that fall below the diagonal more frequently than above it in all forecast systems and combinations, except in the FAS. However, this superiority of the SMM over the single forecast systems is not found in all single aspect of the forecast as noted previously in this study and, for instance, in Hagedorn et al. (2005). As mentioned above, if only the Niño3.4 index is considered, then S4 ensemble-mean predictions are more skilful than the SMM (Figure 4.2). Moreover, if only the target months of July and August for the predictions of the WTI index are considered, then the statistical model and the CFSv2 are more skilful than the SMM.

It is interesting to note that the SMM fails to beat all forecast systems when it has correlation smaller than 0.6 (Figure 4.8). In these cases, all forecast systems and combinations, except S4, MF3 and PC1, have higher correlation than the SMM. The combination methods almost never show a negative correlation. This result is significant because it illustrates an important property of appropriate weighting methods: they reduce the risk of providing poor predictions in cases where the single forecast systems have low or negative correlation.

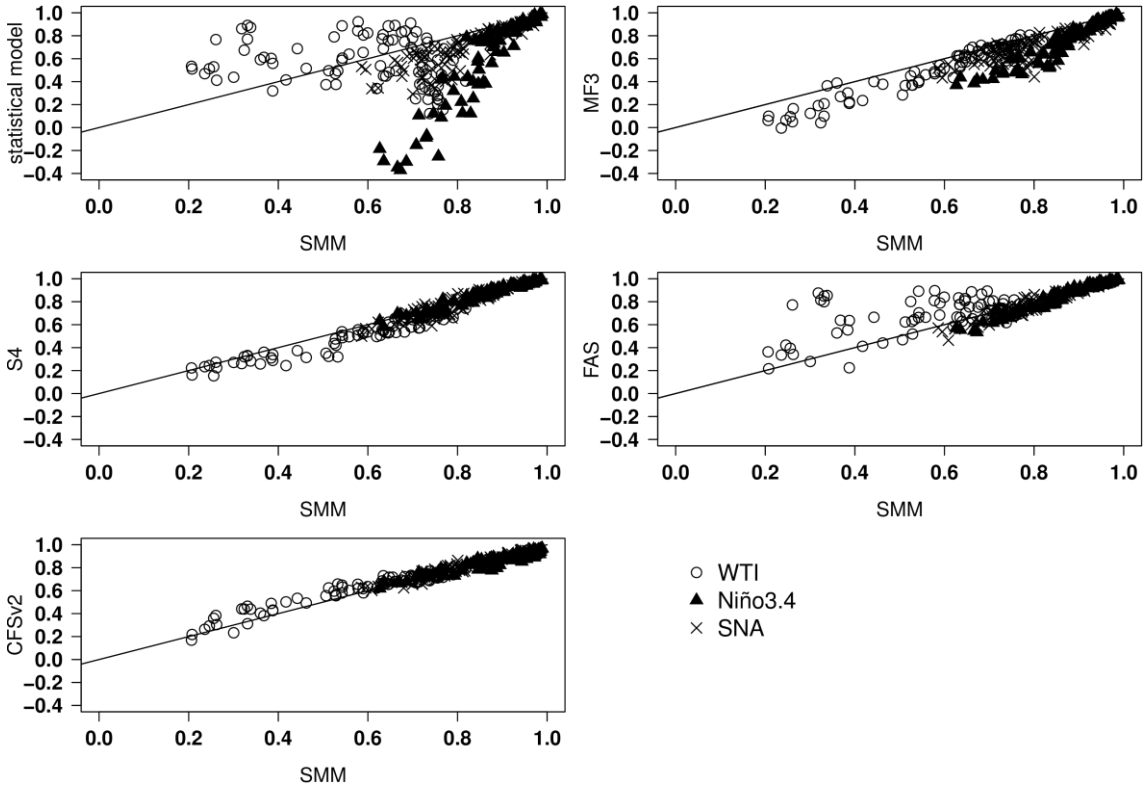

Figure 4.8: Scatterplots of the correlation coefficient for the statistical model, S4, CFSv2, MF3 and FAS versus the SMM. Results are for twelve target months, seven lead times and three indices. Each symbol represents the correlation for one index: WTI (circle), Niño3.4 (triangle) and SNA (cross).

The scatterplots of the BSS show that the SMM has higher skill than the four single forecast systems more frequently than not (Figure 4.9). In agreement with previous studies (Hagedorn et al., 2005) the SMM probabilistic predictions do have an overall improved reliability and resolution when compared to the single forecasting systems (not shown). The statistical model hardly gets values of BSS below -0.4, while the dynamical forecast systems do worse for low values (Figure 4.9). This could be explained because the statistical models are calibrated by construction (Mason and Baddour, 2008) while the three dynamical systems are not and tend to be overconfident (Slingo and Palmer, 2011). None of the weighting methods tends to show higher BSS than the SMM more often than not, but when the SMM has a low BSS the weighted predictions tend to be better.

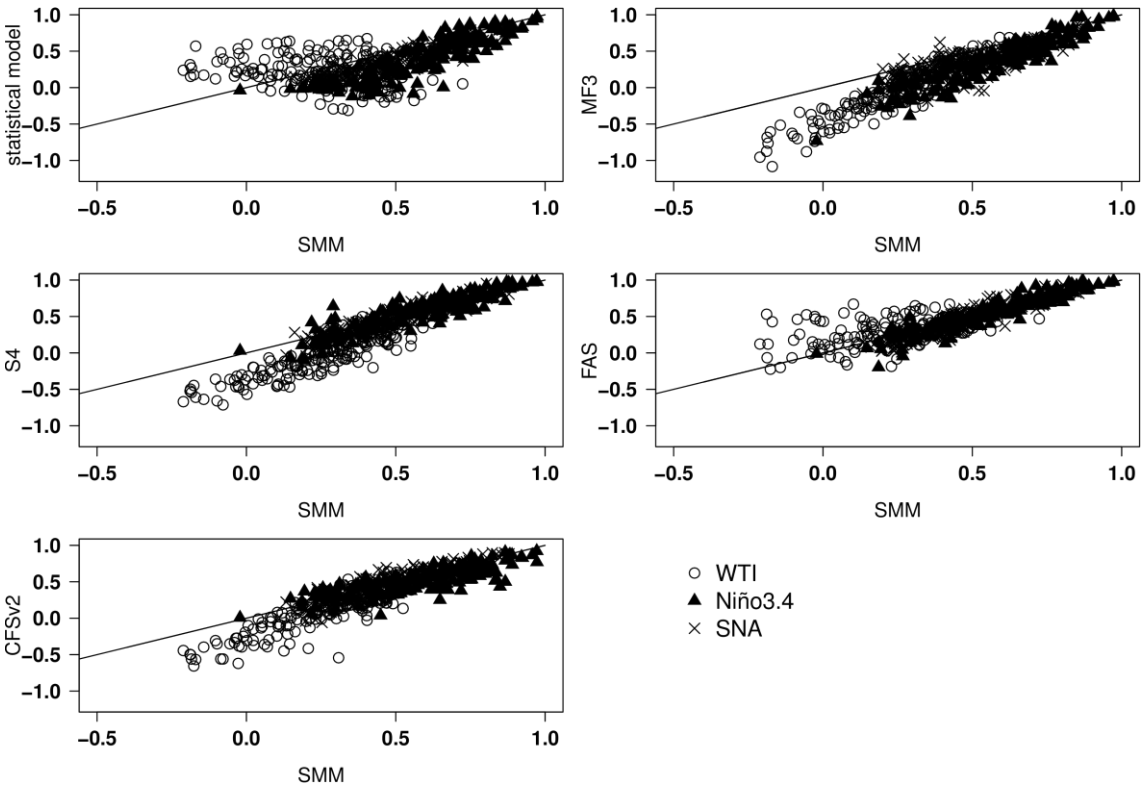

Figure 4.9: As Figure 4.8, but for the BSS. Results also include two events: anomalies above the median and above the upper quartile.

4.3. Summary and conclusions

The chaotic nature of the climate system implies that forecast uncertainty must be quantified (Palmer, 2000). The uncertainties in climate forecasts are due to both the initial conditions and model inadequacy (Slingo and Palmer, 2011). The first source of uncertainty is addressed by generating an ensemble of forecasts, while model inadequacy has been addressed in this thesis using the multimodel method (Doblas-Reyes et al., 2009). It has been shown that the quantification of these two sources of uncertainty leads to more reliable probabilistic forecasts (Coelho et al., 2004; Doblas-Reyes et al., 2005; Hagedorn et al., 2005; Stephenson et al., 2005; Doblas-Reyes et al., 2009; Wang et al., 2009).

Traditionally, multimodel predictions are built by merging the different single forecast systems, avoiding the question of how best to combine them depending on their past performance. Hence, the question of whether there exists an optimal way to combine the different forecast systems remains unanswered. In this chapter, the Bayesian method described in Stephenson et al. (2005) is compared with the multiple linear regression methods described in Doblas-Reyes et al. (2005) and a simple combination method where all forecast systems are combined with equal weight attributed to each of them. However, this study goes a bit farther than those two papers, and several others that were recently published (Palmer et al., 2004; Hagedorn et al., 2005; Kug et al., 2007; Tippett and Barnston, 2008; Kug et al., 2008). In this chapter, the impact of those combination

methods on a series of operational forecast systems, which is an aspect of the problem not dealt with in the past, was investigated. In particular, this implied considering the differences in how the systems are developed in a real-time basis, and how the combination affects predictions that are carried out regularly, with one start date per month. Three operational dynamical seasonal forecast systems were used: S4, CFSv2 and MF3. The statistical model described in Coelho et al. (2004) is used as an additional model for both benchmarking and to increase the number of systems in the combination procedure.

The predictability of the climate system is to a large extent linked to our ability to predict its boundary conditions, such as the SST. Therefore, the forecast quality assessment of the SSTs for deterministic and probabilistic predictions of all forecast systems and combinations was analyzed. Given the large amount of cases considered in this study, for simplicity, the forecast quality assessment is carried out for SST averaged over three different tropical regions: the tropical Pacific Ocean, the tropical Atlantic Ocean and the tropical Indian Ocean.

The SMM, which is used as the reference forecast, has often higher correlation than all single forecast systems. However, for a specific aspect of the forecast the SMM can be outperformed by the best single forecast system as noted in earlier studies (Hagedorn et al., 2005; Wang et al., 2009). For instance, S4 has higher correlation than the SMM more often than not when only the Niño3.4 predictions are considered (Figure 4.2). On the other hand, the statistical model has higher correlation than the SMM more often than not when only the WTI predictions in the boreal summer are considered (Figure 4.6). This shows that empirical systems can still provide useful information. The SMM has also higher mean prediction correlation than all combination methods that assign unequal weights, except for the FAS, more often than not, making it a benchmark difficult to beat. Even the FA method, which includes and generalizes previous calibration methods and had been proved to be competitive against the SMM when predicting the equatorial Pacific SST for the four target months available in the DEMETER project (Stephenson et al., 2005), performed only as good as the SMM in terms of correlation when all cases are considered (all SST indices, target month and lead time pairs). The low number of forecast systems (four) could explain this and short samples (29 years) used to estimate the regression coefficients (Doblas-Reyes et al., 2005).

The SMM outperforms the four single forecast systems analyzed here more often than not also in terms of BSS. It has been found that the higher BSS of the SMM predictions when compared to the single forecast systems is the result of improved reliability and resolution, which is in agreement with previous studies despite the slightly different definition of reliability and resolution used. The probabilistic predictions of the SMM are often better than those of the combination methods that assign unequal weights in terms of BSS. However, some of the results shown here give light to further research on how to improve the SMM predictions using combination methods that assign unequal weights.

For example, FAS deterministic and probabilistic predictions are often competitive against the SMM. The combination methods that assign unequal weights improve the SMM predictions when only a fraction of all single forecast systems have skill as in the case of the SNA index predictions in the boreal fall (Figure 4.5) or in the case WTI index predictions in the boreal summer (Figure 4.7). Therefore, the weighting does not outperform the SMM when the SMM is very skilful, but it reduces the risk of low skill situations that are found when several single forecast systems have a low skill.

5. Prediction of the WAM rainfall regimes

5.1. Introduction

Associated with the apparent motion of the sun, the ITCZ experiences a latitudinal shift along the year that plays a fundamental role in determining the WAM rainfall variability (Motha et al., 1980; Sylla et al., 2013). The WAM rainfall variability spans a wide range of timescales, from intraseasonal (Sultan et al., 2003) to interdecadal (Nicholson, 1993), and is influenced by both local and remote oceanic forcings, and associated changes in the atmospheric circulation (Folland et al., 1986; Fontaine et al., 1995, 1998; Fontaine and Janicot, 1996; Janicot et al., 1998, 2001; Joly and Voldoire, 2009, 2010; Hourdin et al., 2010; Mohino et al., 2011a, 2011b; Rodríguez-Fonseca et al., 2011).

Motha et al. (1980) analyzed long-term rainfall data in Nigeria and found two distinct rainfall patterns. In one of them, rainfall anomalies of opposite signs are observed in the Sahelian and Guinean regions. They suggested that this was associated with the latitudinal migration of the ITCZ such as that above (below) normal rainfall in the Sahelian (Guinean) region is observed when the ITCZ is placed further north of its climatological position. The opposite takes place when the ITCZ does not penetrate into the Sahelian region with its normal intensity. In the second pattern, rainfall anomalies with the same sign are experienced throughout the WAM region.

The two leading modes of WAM rainfall variability, extracted by using PCA, correspond to the rainfall variability along the Sahelian and Guinean regions (Giannini et al., 2003, 2005; Tippet and Giannini, 2006; Philippon et al., 2010). While the Guinean rainfall regime is mostly explained by interannual variations, the variability in the semi-arid Sahelian region occurs mostly on decadal time scales, although interannual variations also play a role in this region, specially linked to ENSO (Fontaine et al., 1998; Janicot et al., 2001; Giannini et al., 2003, 2005; Tippet and Giannini, 2006).

Forecasting the WAM summer rainfall is of great importance, especially taking into account that a large part of this region employs rain fed agriculture (Sylla et al., 2013). On the other hand, farmers find that forecasting the total amount of seasonal rainfall is of limited usefulness (Ingram et al., 2002). Instead, they would benefit from having information such as the duration and distribution of rainfall over time and space or the timing of the monsoon onset (Ingram et al., 2002; Vellinga et al., 2013; Sylla et al., 2013). This kind of information has been hardly taken into account in predictability studies over the WAM region. In this chapter, the seasonal evolution of the WAM summer rainfall is taken into account through the meridional evolution of rainfall from June to October. Latitude-time diagrams of longitudinally averaged rainfall are considered as this approach provides a suitable representation of the integrated atmospheric dynamics of the WAM system, which is related to shifts in the local ITCZ (e.g. Sultan and Janicot, 2000, 2003; Sultan et al., 2003).

AGCMs forced with observed SSTs are able to simulate successfully the two WAM rainfall regimes (Giannini et al., 2003, 2005; Tippet and Giannini, 2006). However, Goddard and Mason (2002) compared the ensemble-mean anomaly correlation simulated and predicted by an AGCM using persisted SST anomalies and found that errors in the predicted SST could lead to a significant degradation of the predictive skill. They showed that the WAM rainfall during the July-August season is one of the most severe examples of this loss of prediction skill. In a different study, Tompkins and Feudale (2010) noticed that a dipole bias in the WAM rainfall prediction by S3, with dry (wet) conditions over the Sahel (Gulf of Guinea). A warm bias in the Equatorial Atlantic SST predictions by S3 would affect the observed northward migration of the ITCZ. When S3 is run with observed SST as boundary forcing the dipole bias disappears, and an overall reduction in rainfall bias is found (Tompkins and Feudale, 2010). Capture the interannual variability of the Equatorial Atlantic SST using simulations from the Coupled Model Intercomparison Project 3 (CMIP3) is still an issue; and consequently, its influence on the rainfall over the Western African continent is hardly reproduced (Joly and Voldoire, 2010).

Cook and Vizy (2006) studied the ability of eighteen climate models to simulate the climatology and the dipole mode of WAM variability associated with the meridional migration of the ITCZ. They found that all of them have positive SST bias in the Gulf of Guinea, only ten could simulate the main observed climatological features (e.g., some of the forecast system put the maximum rainfall over the ocean due to the warm SST biases) and only eight the dipole mode of variability. An analysis of the recently available CMIP5 historical simulations shows that dynamical forecast systems still have substantial SST biases in the Equatorial Atlantic (Roehrig et al., 2013). Zuo et al. (2013) used the CFSv2, to assess the predictability of the modes of interannual rainfall variability of three northern hemisphere monsoon systems: the Asian and Indo-Pacific, the West African and the North American monsoon systems. They found that the low predictability of the PCs associated with the two main modes of the WAM rainfall variability could be probably due to the link between the WAM and the equatorial Atlantic SST, which is poorly predicted by the CFSv2.

In addition, predictive skill can be negatively affected if the model used to take advantage of SST information does not properly describe the mechanisms responsible for the WAM rainfall (Kumar et al., 2005). Im et al. (2014) used a regional climate model with observations and reanalysis as initial and boundary conditions to show the sensitivity of the WAM rainfall, surface energy balance and circulation to the land surface and convection schemes. They show that predictability of these parameters over the WAM can be significantly improved when the land surface and convection are better represented in the model. Zuo et al. (2013) found that the poor representation of land surface processes in the CFSv2 could in part explain the low predictability of this forecast system when predicting the WAM rainfall regimes. Improving the representation of land surface and rainfall processes in dynamical forecast system is difficult and the skill improvement in

one region is usually followed by a degradation in another one so that the overall improvement is usually small (Tompkins and Feudale, 2010).

When systematic errors are important, several studies have shown how the combination of several dynamical forecast systems yields on average better deterministic and probabilistic forecast skill than any of the single systems (Coelho et al., 2004; Doblas-Reyes et al., 2005; Hagedorn et al., 2005; Stephenson et al., 2005; Batté and Dequé, 2011; Rodrigues et al., 2014a). It has been shown that combining statistical and dynamical forecast systems could enhance forecast skill even further (Coelho et al., 2004; Stephenson et al., 2005). Coelho et al. (2004) and Stephenson et al. (2005) used the FA technique, a Bayesian method for calibrating and combining several dynamical forecast systems taking into account historical (observed) information, to forecast SST over the Pacific region. The FA technique assigns weights to each forecast system in the combination procedure based on each systems' forecast error (i.e. more weight to forecast systems with less forecast error). Stephenson et al. (2005) found that the FA technique could improve forecasts not only over the single systems but also over the SMM combination, where all forecast systems are combined assigning equal weights. Rodrigues et al. (2014a) studied the benefits of combing three operational dynamical forecast systems and a simple statistical model to predict SST over three ocean basins. They found that on average the SMM is better than the single forecast systems and the combination methods that assign weight to each forecast system, including the FA. On the other hand, assigning different weights could reduce low skill when most forecast systems perform badly, which is typically the case for the WAM precipitation.

Previous multimodel assessments, however, showed limited benefit of merging different sources of information. Bouali et al. (2008) found that the DEMETER multimodel system has only modest skill when predicting the Sahelian rainfall. Philippon et al. (2010) studied the skill of the ENSEMBLES stream 1 multimodel when forecasting key parameters of the WAM and found that the Guinean rainfall regime could be accurately predicted by these systems, but not the Sahelian regime. Batté and Dequé (2011) used the ENSEMBLES stream 2 forecast systems to study the precipitation seasonal forecast skill over Africa and found that the SMM improves on average forecast skill over the single systems. They also found that probabilistic forecasts were more skilful in the Guinean region than in the Sahelian region. Vellinga et al. (2013) used several forecast systems, including the ones from the ENSEMBLES project and the UK Met Office operational seasonal forecast system UK Met Office global seasonal forecast system 4 (GloSea4), to study the skill of these systems when forecasting the onset of the WAM rainy season. They found that these forecast systems have modest probabilistic skill when forecasting the onset of the Sahelian rainfall. This was attributed to the difficulty of such systems to capture the mean rainfall amount in the Sahel and the influence of a diversity of intraseasonal phenomena that usually have little or no predictability at this timescale.

New aspects of seasonal climate prediction of the WAM are addressed in this chapter. Firstly, a targeted methodology to assess both the seasonal evolution of the WAM rainfall within a rainy season and its interannual variability simultaneously is considered. Secondly, the two leading modes of the WAM rainfall variability are estimated using the seasonal evolution diagrams over the whole hindcast period. The robustness of the methodology was estimated using two different datasets to assess the uncertainty associated with the observations. Thirdly, several quasi-operational forecast systems were used to estimate the leading modes of WAM rainfall variability. The aim is to assess the ability of the forecast systems to predict the seasonal evolution of the latitudinal migration of rainfall over West Africa. A simple statistical model that uses SST indices as predictors for the WAM rainfall regimes is considered as both a benchmark and an additional forecast system. Fourthly, three methods of combination described in Chapter 3 (i.e. SMM, FAS and FAC) are used to combine the dynamical and empirical seasonal predictions. Fifthly, a forecast quality assessment of the combinations and the single forecast systems is performed using multiple deterministic and probabilistic verification measures.

5.2. Forecast quality assessment

Figure 5.1 shows the correlation between the predicted ensemble mean and the observed JAS rainfall at each grid point over the WAM region for the period 1982-2011. The correlation is computed for all dynamical forecast systems at lead time 2 months (i.e. predictions starting in May). The aim is to assess the ability of these systems to predict the spatial distribution of the seasonal WAM rainfall. S4 shows positive correlations in almost all grid points at the three start dates analyzed (results for the two start dates of June and April are not shown), most of them with p-values smaller than 0.10. On the other hand, CFSv2 has low, and in several instances, negative correlation values. Most of the positive correlation values in this forecast system appear north of 10ºN, over the Sahel. Correlation values below 0.1 are found more often in the region south of 10ºN. MF3 also has low correlation skill when compared to S4, but contrary to CFSv2, most of the positive correlation appears south of 10ºN in the Guinean region over the longitudinal range 20ºW-10ºE. CCSM3 performs generally worse than the previous three forecast systems but, as MF3, it performs better over the Guinean region. GFDL performs well in almost all grid points and lead times, except for the Guinean region at lead time 3 (i.e. predictions starting in April), where it has correlation values below 0.1 more often than above it (not shown). As in the CFSv2 case, GFDL performs better over the Sahel than over the Guinean region. The IRI-ECHAM systems perform poorly over the WAM region, especially over the Sahel where they have negative values more often than positive ones. CMC2 shows positive correlations all over the WAM region at the three start dates analyzed (i.e., lead times 1, 2 and 3 months).

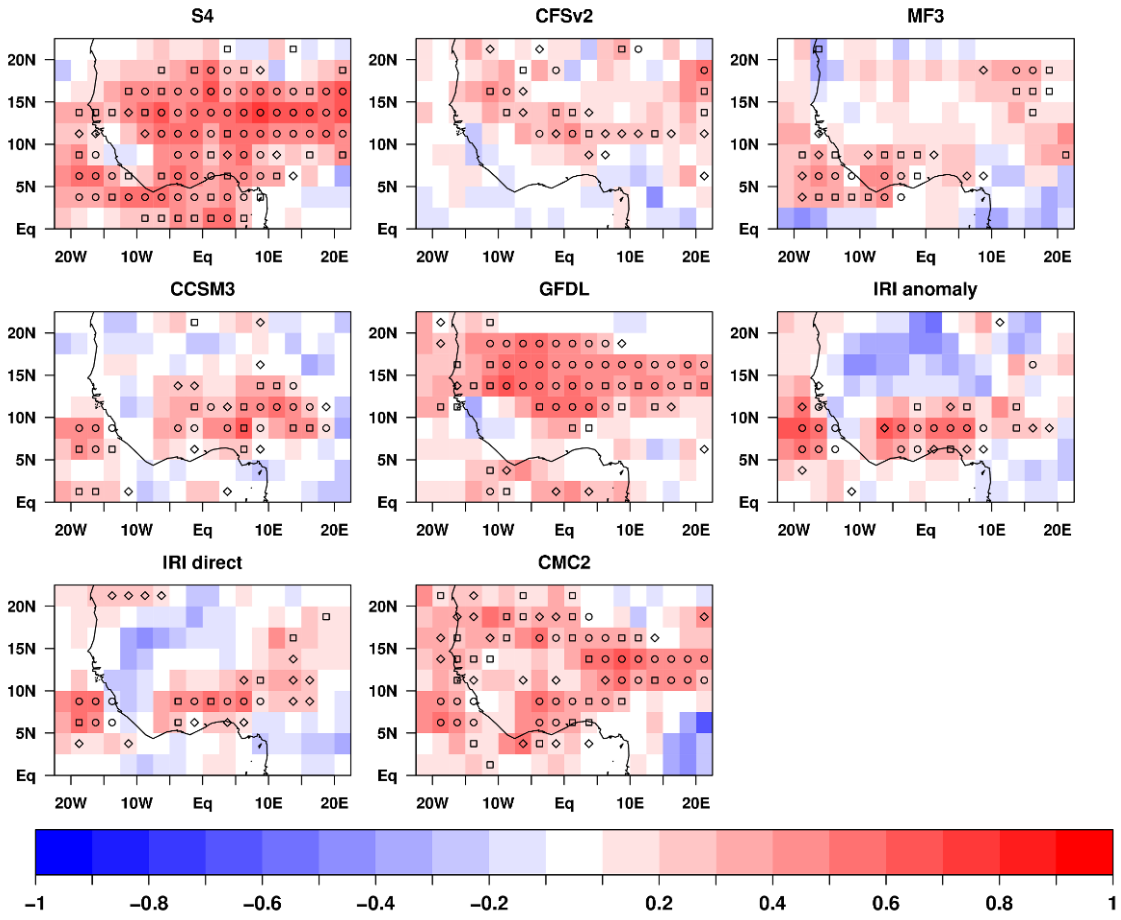

Figure 5.1: Correlation coefficient between the predicted ensemble mean and observed summer (JAS) rainfall at each grid-point over the WAM region for the period 1982-2011. The GPCP dataset was used as the reference data. Forecasts are for lead time 1 month and interpolated into the GPCP grid prior to computing the correlation coefficient. Circles are for p-values smaller than or equal 0.01, squares for p values between 0.05 and 0.01, and diamonds for p values between 0.10 and 0.05.

Figure 5.1 illustrates that S4 has the highest overall correlation skill at all lead times, followed by CMC2 and GFDL, respectively. This is a feature that will appear in many other of the diagnostics described in this chapter. S4 seems to represent a leap forward in the seasonal prediction of the WAM precipitation with respect to previous versions of this system and to other contemporaneous operational systems. This leap forward can be measured when compared with the performance of the previous ECMWF forecast system, which had similar skill to other European systems (Batté and Déqué, 2011). The grid-point correlation over the WAM region does not substantially change with lead time in any of the forecast systems.

S4, GFDL and CMC2 have smaller mean systematic errors when compared to the other systems (not shown). Therefore, even though a direct link between mean biases and forecast skill could not be established, one can expect that improving the physical processes that are at the origin of the model drift and the systematic error, and that hamper

the conversion of predictability into skill, could lead to improvements in forecast quality (DeWitt, 2005). Molteni et al. (2011) found that S4 has improved the simulation and prediction of ocean/atmosphere variability in the tropical Atlantic and adjacent regions when compared to S3, which could benefit the prediction of the WAM rainfall. They highlight that some of the improvements S4 has achieved when compared to its predecessor might be due to higher horizontal and vertical resolution, a more accurate initialization of the land-surface variables, and improved physical parameterization, among other reasons. In fact, it is observed that the higher the model resolution of a system is, the smaller its biases are, with CCSM3 being the only exception to this simple rule. However, as pointed out by Kirtman et al. (2014), CCSM3 is generally worse when compared to the other NMME systems in terms of root mean square error (RMSE) of the tropical SST for September start dates at leads 0-5 months. As a consequence, it is planned to be replaced by CCSM4 in the second phase of the NMME project (Kirtman et al., 2014). On the other hand, not always a small bias leads to a high correlation. For instance, CFSv2 shows a relatively small bias, while at the same time it has low correlation.

The WAM rainfall displays a strong monthly variability, which is illustrated by considering the latitudinal migration of the zonally-averaged rainfall between the months of June and October. Figure 5.2 shows the climatology of monthly-mean rainfall averaged over 10ºW- 10ºE and displayed over the latitudes between the Equator and 20ºN. The climatology is computed using the GPCP dataset for the period 1982-2011 (upper left panel), GPCP after applying a mask over the ocean for the same period (upper right panel), GPCC for the same period (lower right panel) and the GPCC dataset for the period 1951-2011 (lower left panel). The climatologies of the zonally averaged monthly rainfall have similar patterns in both GPCP and GPCC. They show a northward migration of the rainfall that reaches its northernmost position at 18ºN in July and August, moving southward later in the year. Some differences between GPCC and GPCP can be found over the common period. These differences already point at the observational uncertainty of the WAM rainfall.

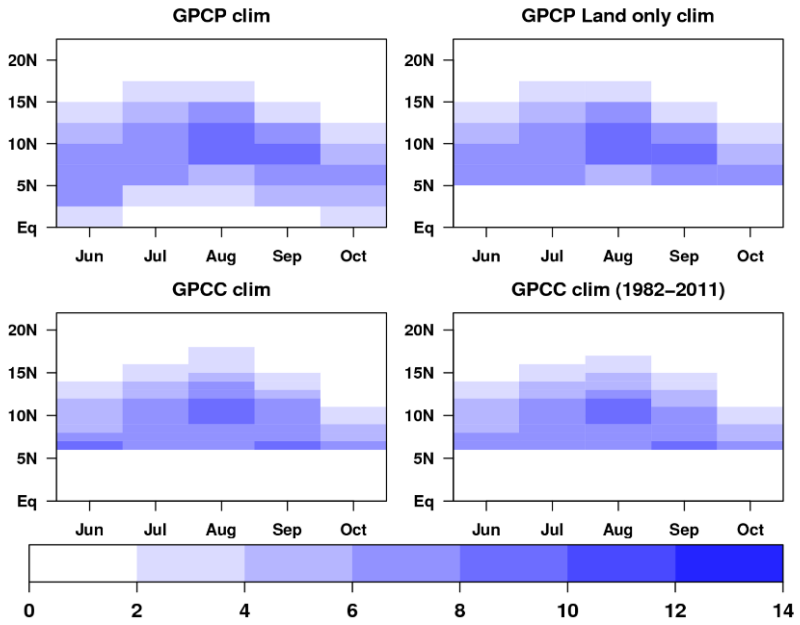

Figure 5.2: Monthly rainfall (mm/day) averaged over 10ºW-10ºE as a function of month, from June to October, and of latitude. Climatologies of the two analysed observational datasets, GPCP and GPCC were computed using the period 1982-2011 and 1951-2011, respectively, except when indicate otherwise. For comparison, the GPCP climatology was also computed masking the ocean and the GPCC using only the common period 1982-2011.

Every dynamical forecast system successfully simulates the meridional shift of the rainfall for the three lead times analyzed (not shown). However, they all fail in simulating the correct position and magnitude of the rainfall maxima and therefore have substantial biases, suggesting that these systems do not fully reproduce the physical processes associated with the WAM rainfall. As an illustration, Figure 5.3 shows the systematic error of the dynamical forecast systems at lead time 1. CCSM3 has a larger bias than the other forecast systems. It not only fails to simulate the rainfall maxima in August but it is the only forecast system that simulates rainfall above 2 mm/day north of 18ºN. MF3 also has substantial biases. In particular, it has a positive bias south of 10ºN and negative north of it indicating that in this forecast system the ITCZ does not penetrate as far north as in the observations, which creates a dipole-like bias pattern (i.e. excessive precipitation at lower latitudes and a deficit at higher latitudes). This pattern is also observed in S4 and CFSv2 but with smaller magnitude when compared to MF3. The IRI-ECHAM systems and CMC2 have a dipole-like pattern with inverse sign (i.e. excessive precipitation at higher latitudes), while GFDL has a positive bias overall. The forecast systems could be ranked in decreasing order of the mean bias (i.e. sum of the mean bias over the whole domain) at lead time 1 to give S4, CMC2, CFSv2, IRI-ECHAM direct, GFDL, MF3, IRI-ECHAM ano and CCSM3. As it was also found in the analysis without the longitudinal averaging (not shown), the systems with lower (higher) systematic errors are the systems with higher (lower) resolution, CCSM3 being the only exception to this.

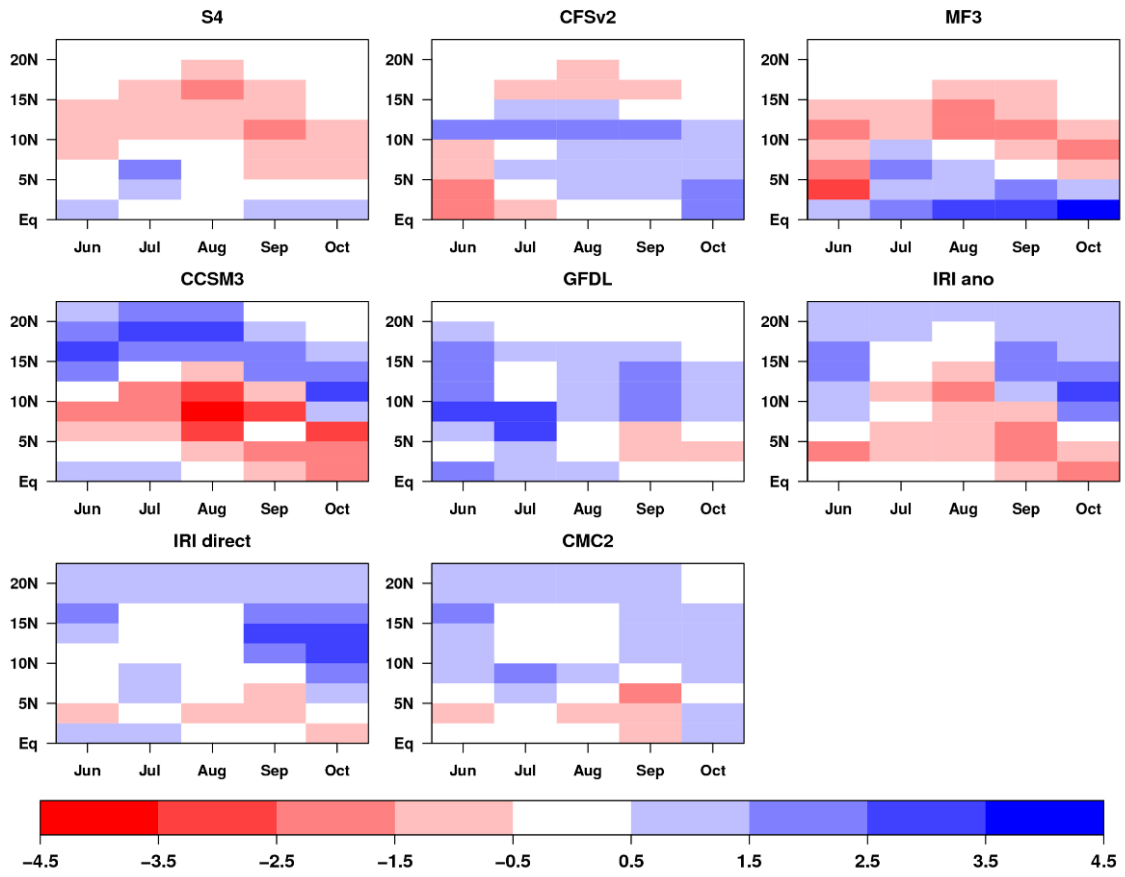

Figure 5.3: Mean precipitation bias (mm/day) of the dynamical forecast systems over the WAM region for the period 1982-2011 is computed as the difference between the one-month-lead hindcasts and the GPCP mean climatological estimates. The hindcasts were interpolated into the GPCP grid prior to computing the systematic error.

The two leading modes of the observational WAM rainfall, obtained with the PCA method described in Chapter 3, are shown in Figure 5.4. The aim of the longitudinal averaging applied to the data prior to the PCA is to concentrate in both the latitudinal migration and the seasonal distribution of the WAM rainfall. The first EOF (EOF1) in the GPCP dataset shows positive values south of 10ºN, in the Guinean region, while the second EOF (EOF2) shows positive values north of 10ºN, in the Sahelian region. The variance associated with these two EOFs is 29% and 23%, respectively (Table 5.1). This is in agreement with the WAM patterns described in the literature using different methodologies (Motha et al., 1980; Fontaine et al., 1995; Fontaine and Janicot, 1996; Janicot et al., 1998; Giannini et al., 2003, 2005; Mohino et al., 2011b; Rodríguez-Fonseca et al., 2011). The same analysis has been performed on the GPCP dataset after applying a mask over the ocean and on the GPCC dataset with a common period 1982-2011 and an extended period 1951-2011 to assess the observational uncertainty. The GPCP land-only and the GPCC datasets have a reverse order of the leading modes when compared to the GPCP land-ocean (Figure 5.4). This reverse pattern when land-ocean and land-only data are used has been documented previously (Giannini et al., 2005). This reversal is probably due to the variance maximization of inland rainfall, where the latitudinal migration from the ocean into the Guinean region early in the season is not considered.

The variance explained by these two EOFs varies, being 31% (EOF1) and 24% (EOF2) in the GPCP land-only, 27% and 20% in the GPCC, and 30% and 18% in the GPCC for the common period 1982-2011. The difference between the smallest and largest values of the variance explained in the observational datasets is 4% and 6% for the first and second EOF, respectively. As previously with the mean bias, this uncertainty in the observations will be taken into account when interpreting the EOFs from the hindcasts.

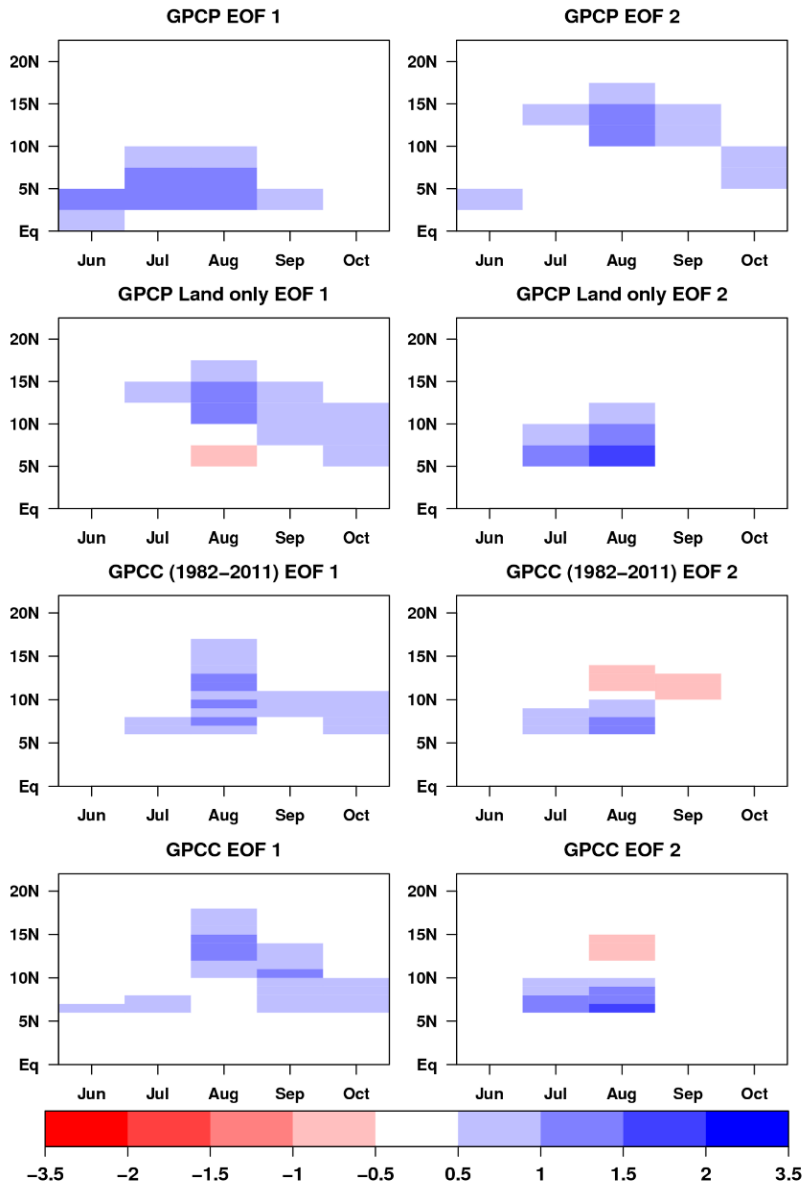

Figure 5.4: Leading two EOFs of the longitudinally-averaged precipitation datasets of Figure 5.2.

Table 5.1: Variance explained (%) by the first and second modes of the WAM rainfall variability by the GPCP, GPCC, and the dynamical forecast systems. For the predicted modes of variability, the variance is displayed for each lead time.

	Variance (%): First mode			Variance (%): Second mode		
	Lead 0	Lead 1	Lead 2	Lead 0	Lead 1	Lead 2
GPCP		29			23	
GPCP land-only		31			24	
GPCC		27			20	
GPCC (1951-2011)		30			18	
S4	25	34	41	15	14	11
CFSv2	15	19	18	09	09	08
MF3	27	20	16	11	11	11
CCSM3	46	49	51	10	09	09
GFDL	24	22	30	19	18	18
IRI-ECHAM ano	34	31	29	14	15	15
IRI-ECHAM direct	32	33	31	11	12	11
CMC2	18	18	15	12	12	13

To illustrate that the Guinean regime is captured in the EOF1 (EOF2) when the dataset have values over both land and ocean (land only) and vice-versa for the Sahelian regime, the PCs associated with these EOFs are displayed in Figure 5.5. The first PC (PC1) of the GPCP dataset is highly correlated with the second PC (PC2) of the GPCP land-only and GPCC datasets, and vice versa (see figure caption). The GPCC PCs show that the Guinean regime is characterized mainly by interannual variability while the Sahelian regime is associated with substantial interdecadal variations, although interannual variations also play a role in the latter as described in previous studies (Fontaine et al., 1998; Giannini et al., 2003, 2005). The PCA has been also performed over the full spatial field (i.e. without longitudinal averaging) of the GPCP JAS rainfall with longitudes 10ºW-10ºE and latitudes between the Equator and 20ºN. The aim is to compare the modes of variability of the WAM rainfall computed in a conventional way by applying PCA on the seasonally-averaged spatial field with the ones computed by applying the PCA on the longitudinally-averaged seasonal evolution diagrams shown above. The first and second EOFs of the JAS full spatial field are also associated with the Guinean and Sahelian regimes, respectively (not shown). The lower panels of Figure 5.5 show the PCs associated with the Guinean and Sahelian regimes estimated by applying the PCA on both the full spatial field and the seasonal evolution diagram. As expected, the PCs are highly correlated in both cases, being the correlation 0.91 for the Guinean regime and 0.90 for the Sahelian regime. Even so, the zonally-averaged rainfall approach allows a better characterization of the intraseasonal evolution of the rainfall regimes because the rainy seasons associated with the two modes are not simultaneous.

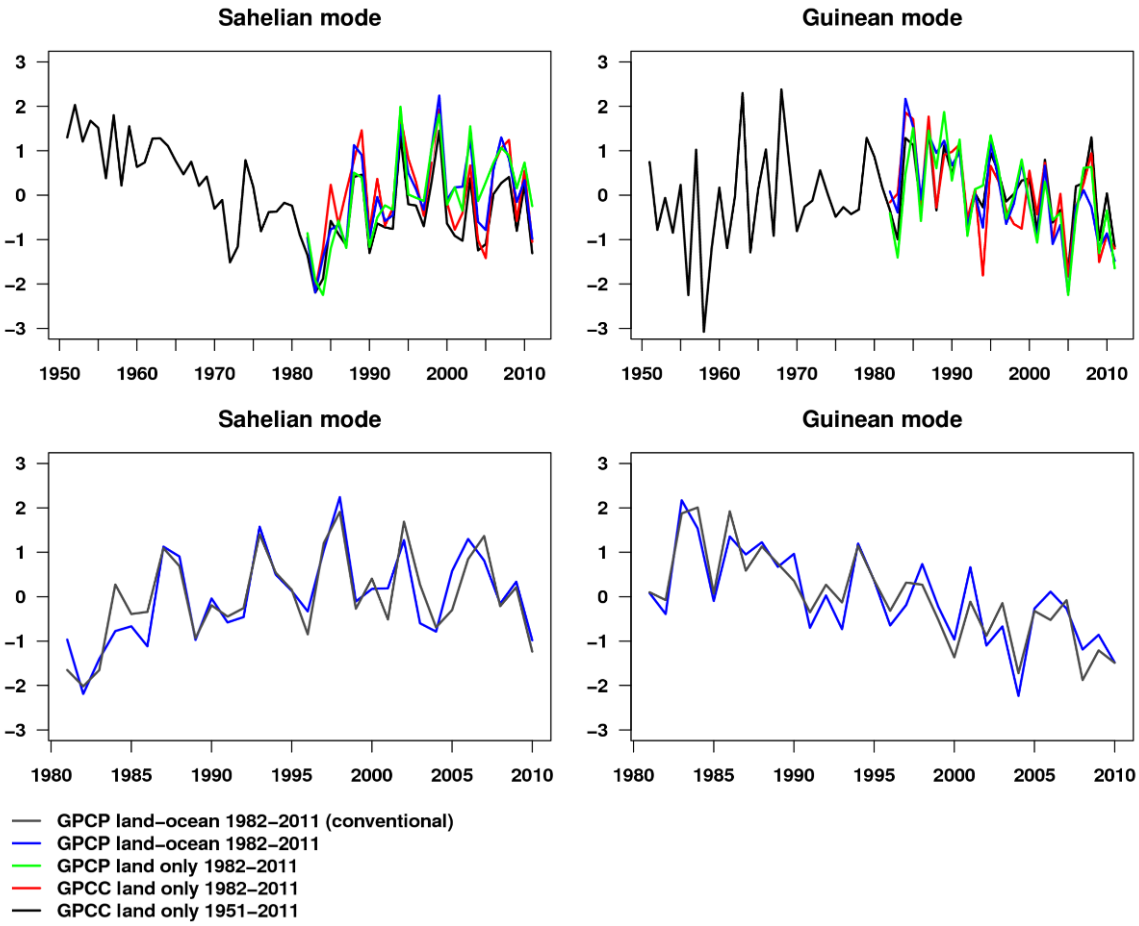

Figure 5.5: Principal components associated with the EOFs shown in Figure 5.4. The blue line is the PC of the GPCP land and ocean, the green line is the PC of the GPCP land only and the red line is the PC of the GPCC land only. These three PC are computed for the common period 1982-2011. The black line is the PC computed using the GPCC land only for the period 1951-2011. These PCs are estimated using the seasonal evolution diagrams averaged over 10ºW-10ºE, covering the latitudes between the Equator and 20ºN and the period between June and October. For comparison, the PCs are also estimated using the traditional way with the full spatial field (i.e., without applying the longitudinal averaging) over 10ºW-10ºE and between the Equator and 20ºN on the JAS rainfall (gray line, bottom panels). The blue lines are the same in the top and bottom panels. The correlation between GPCP land and ocean PC1 (blue line in the upper right panel) and the GPCC land only PC2 (black line in the upper right panel) is 0.84 while the correlation between the GPCP land and ocean PC2 (blue line in the upper left panel) and the GPCC land only PC1 (black line in the upper left panel) is 0.95. The correlation between the WAM rainfall regimes estimated using the seasonal evolution diagrams and the spatial field is 0.91 for the Guinean regime (lower right panel) and 0.90 for the Sahelian regime (lower left panel).

The first EOF of the dynamical forecast systems reproduces the overall features associated with the observed Guinean regime, as they locate the positive values south of 10ºN and capture the northward migration of the rainfall (Figure 5.6 illustrates the results for lead time 1 month). This is similar to what is found in the GPCP land and ocean

dataset. However, the forecast systems fail to simulate the accurate magnitude and location of the maxima of the observed EOF, and some of the forecast systems even reproduce a pattern different to the one found for the observations in Figure 5.4. S4's EOF1 closely resembles the GPCP EOF1 pattern. The variance explained by S4's EOF1 varies considerably with lead time, from 25% at lead time 0 (underestimated when compared to GPCP) to 34% and 41% at lead times 1 and 2 months, an important overestimation when compared to all the observational estimates (Table 5.1). This could be explained by the fact that S4 underestimates (overestimates) the Guinean rainfall at lead time 0 (2) months with respect to GPCP. This is likely due to the increasing SST bias with forecast time in the Equatorial Atlantic (Doblas-Reyes et al., 2013a). CFSv2 also captures well the Guinean regime's pattern, albeit overestimates the role of the rainfall in September and October. MF3 captures the anomalous rainfall in June, July and August as in the GPCP dataset, but overestimates it in several latitudes and target months. Surprisingly, despite its large systematic errors (Figure 5.3), CCSM3 captures the rainfall evolution anomaly in June, July and August, but overestimates the duration of the anomalous rainy season. In addition, CCSM3 overestimates the variance explained by the EOF1 at all lead times and has the largest difference when compared to GPCP (Table 5.1). GFDL generally overestimates the rainfall anomalies, but differently from the previous forecast systems it yields rainfall above 10ºN and in several latitudes in the target months of September and October. Both IRI systems place the rainfall maximum in June and thus, overestimate the rainfall at this target month. IRI-ECHAM anomaly underestimates the observed rainfall anomaly maxima in July and August and overestimates the rainfall latitudinal extent later in the season, while IRI-ECHAM direct simulates better than IRI-ECHAM anomaly the rainfall maxima, but overestimates the signal in September and October. The IRI-ECHAM systems overestimate the variance explained by the first EOF at all lead times, except for the IRI-ECHAM anomaly at lead time 2 (Table 5.1). CMC2 generally underestimates the amplitude of the pattern, although it shifts the pattern north of 10ºN, contrary to what is found in GPCP. CFSv2, MF3 and CMC2 underestimate the variance explained by the first EOF at all lead times.

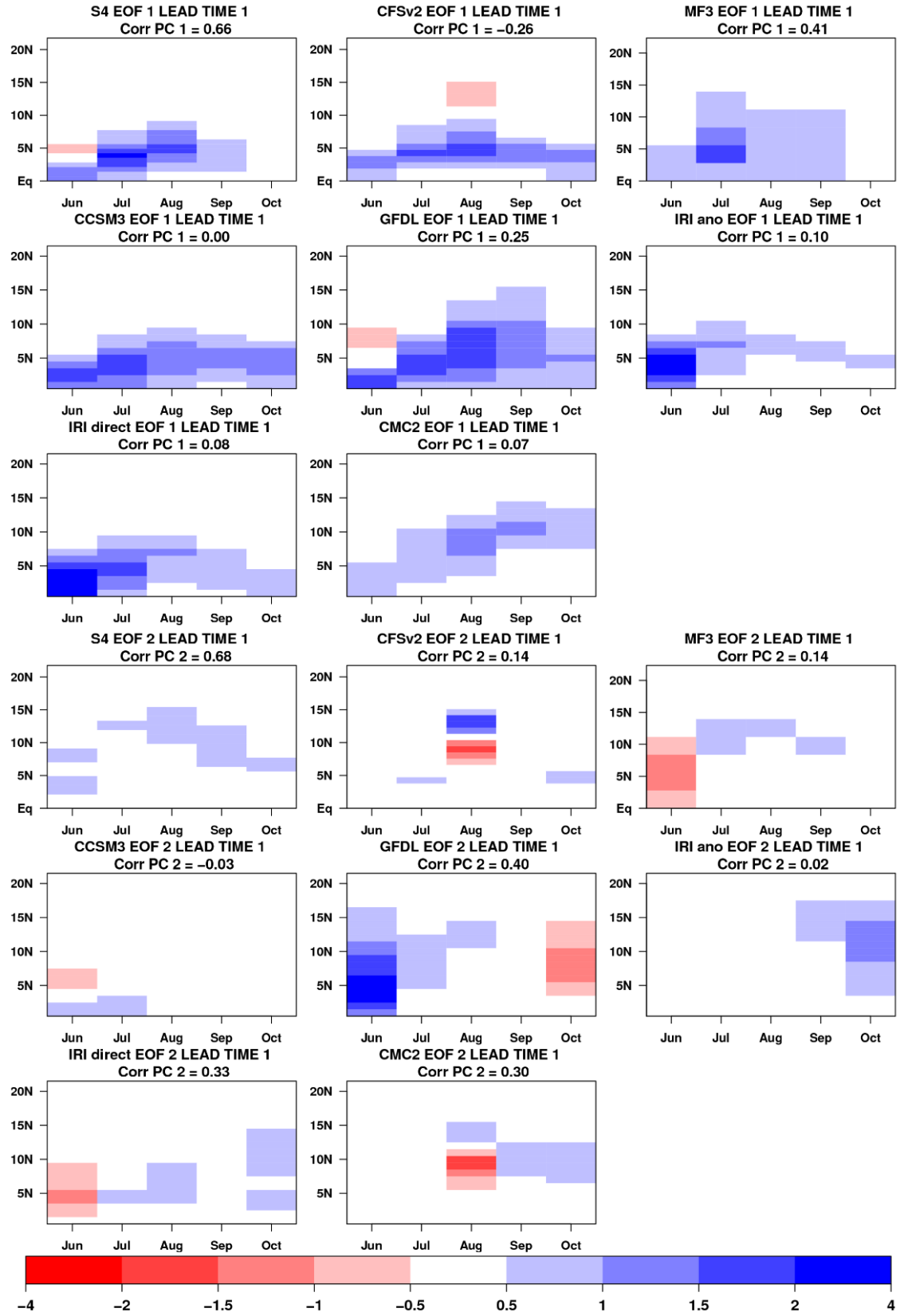

Figure 5.6: As Figure 5.4 but for the lead time 1 month (start date in May) dynamical hindcasts. EOF1 is displayed in the upper set of panels and EOF2 in the lower set of panels. The correlation between the predicted and observed PCs is included in the second line of the panel title.

Contrary to the Guinean regime, the Sahelian regime is only well simulated by S4, yet the amplitude of the pattern is generally underestimated when compared to GPCP. CFSv2 captures the pattern north of 10ºN, but gives an unrealistic pattern with a signal of opposite sign south of 10ºN in August. MF3 also captures the pattern north of 10ºN in July and August, but has a pattern of opposite sign in June. CCSM3 completely fails to simulate any signal north of 10ºN. GFDL captures the pattern in the Sahelian region in August, but shows a pattern of similar sign in June and of opposite sign in October, which are not found in the GPCP pattern. Both IRI-ECHAM systems completely fail to capture the Sahelian regime. CMC2 captures the Sahelian signal but, as other systems do, also simulates a pattern of opposite sign south of 10ºN. All forecast systems underestimate the variance explained by the second EOF when compared to GPCP EOF2 at the three lead times (Table 5.1), which is supposed to be related to the problems all the systems have to timely shift the precipitation over the Sahel during the rainy season.

Figure 5.7 illustrates the indices for the Guinean rainfall regime predicted by the statistical model, the dynamical forecast systems and their combinations. The predictions shown are for lead time 1 month (i.e. predictions starting in May). Several deterministic and probabilistic scores are also displayed. The zero line is shown for reference. The statistical model, which is based on the May Atl3 index as predictor, captures well the interannual variability associated with the Guinean regime. The correlation coefficient of the statistical model is the third largest among the single forecast systems, being outperformed only by S4 and MF3 and it is one of the few systems that has a positive BSS. In addition, the statistical model outperforms all forecast systems and combinations in terms of reliability skill score when predicting the Guinean rainfall above the median at lead times 1 (Figure 5.7) and 2 months (not shown). This illustrates that simple linear regression models are still difficult to be beaten by state-of-the-art dynamical forecast systems, especially in the tropical Atlantic basin.

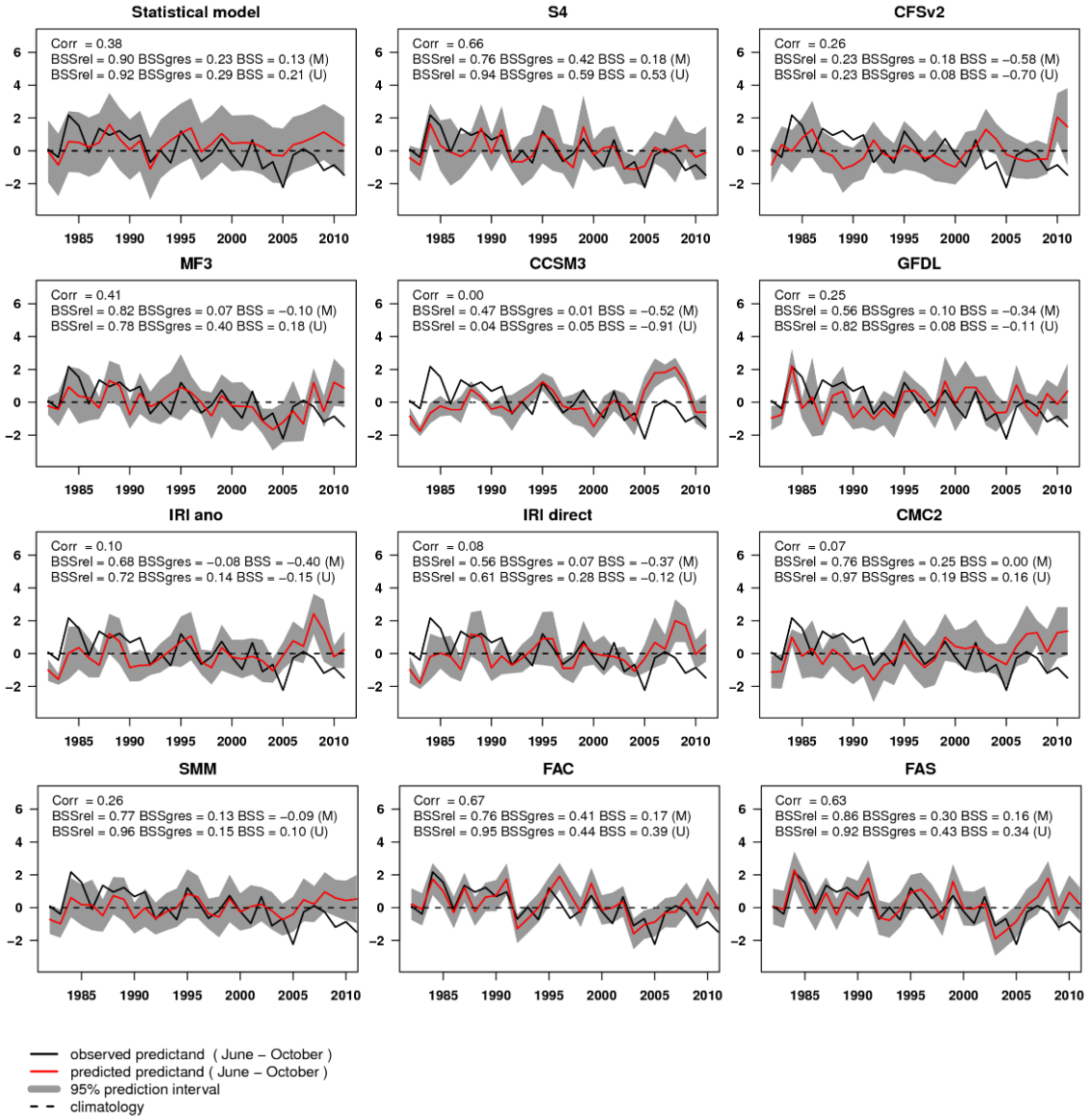

Figure 5.7: Leading principal component (Guinean regime) predicted by the statistical model, the dynamical forecast systems and their combinations. Predictions are for lead time 1 (start date in May). Observed values (black solid line), predicted values (red solid line), 95% predicted interval (grey area) and the zero line (black dashed line) are displayed. The values displayed are anomalies. The correlation coefficient, the BSS, BSSrel and BSSgres for probabilities of rainfall regime being above the median (M) and the upper quartile (U) are displayed in each panel.

Following on its excellent representation of the Guinean rainfall spatial-temporal pattern, S4 captures the interannual variability associated with the Guinean regime and its ensemble-mean correlation is 0.66. S4 is also skillful probabilistically, with most of the observations falling inside the 95% predicted interval (a sign of reliability). The resulting positive BSS values are among the three largest for the two binary events described in this study. Additionally, in most cases it shows the best resolution skill score for the Guinean regime above the median and upper quartile at the three lead times. MF3 has

lower skill than S4, but still shows a high ensemble-mean correlation, while the BSS ranges between negative values (event above the median) and low positive ones (0.18 for the event above the upper quartile). CFSv2 and GFDL have positive correlation of 0.26 and 0.25, respectively, but no positive skill in terms of BSS. Finally, CCSM3, the IRI-ECHAM systems and CMC2 have no deterministic or probabilistic skill when predicting the Guinean regime with 1 month lead time.

It has been considered difficult to improve the SMM forecasts using combination methods that assign different weights to the forecast systems based on the past performance (DelSole et al., 2012). In this case, when the different forecast systems are brought together, the SMM performs worse than the weighted combinations FAC and FAS. This can be explained because weighting methods can provide more skillful forecasts than the SMM when most systems perform badly and there is a small subset that stands out (Rodrigues et al., 2014a). When comparing with all the forecast systems available, the FAC has the best correlation coefficient (Figure 5.7), which is slightly higher than S4 and FAS. On the other hand, both FAC and FAS are outperformed by S4 in terms of BSS, reflecting the difficulty that combination methods have to conserve the forecast resolution when producing more reliable predictions.

A summary of the forecast quality measures for both the Guinean and Sahelian regimes and the three lead times considered can be found in Figure 5.8. The statistical model has only one correlation value for each WAM regime (the correlation does not vary with lead time) as it takes advantage of using the best SST predictor for each regime (see Appendix B for detailed information). Interestingly, a statistical model based on simple linear regression still provides useful information and beats most of dynamical forecast systems when predicting the Guinean and the Sahelian regime. Only S4 and MF3 outperform the statistical model when predicting the Guinean regime, and S4 and GFDL (for lead time 0), S4, GFDL, IRI-ECHAM direct, and CMC2 (for lead time 1) and only S4 (for lead time 2) when predicting the Sahelian regime.

S4 has the highest correlation when predicting both rainfall regimes at all lead times, with two exceptions in the prediction of the Guinean regime: FAS is the best at lead time 0, while FAC is the best at lead time 1 (Figure 5.8). As mentioned above, S4 has improved when compared to its predecessor when predicting the WAM variability (Molteni et al., 2011). S4 has correlation above 0.6 in all cases, except for the Guinean regime at lead time 2 months. Interestingly, the S4 correlation for the Sahelian regime does not vary much with lead time. MF3 (GFDL) is only competitive when predicting the Guinean (Sahelian) regime with average correlation of about 0.45. On the other hand, CFSv2 has no skill when predicting the Guinean regime and low correlation when predicting the Sahelian regime. CCSM3, CMC2 and both IRI-ECHAM systems perform generally worse than the other dynamical forecast systems. As pointed out previously, the SMM usually outperforms unequal methods of combination when most single forecast systems have skill, as in the Sahelian regime at lead time 1. The opposite would happen when only

a fraction of the forecast systems have skill as in most cases in Figure 5.8. However, in this study, S4 is an outlier when predicting the WAM rainfall variability modes as this system is far better than any other single forecast system. Therefore, combining it with the other forecast systems will hardly improve the forecast quality of the WAM rainfall regimes.

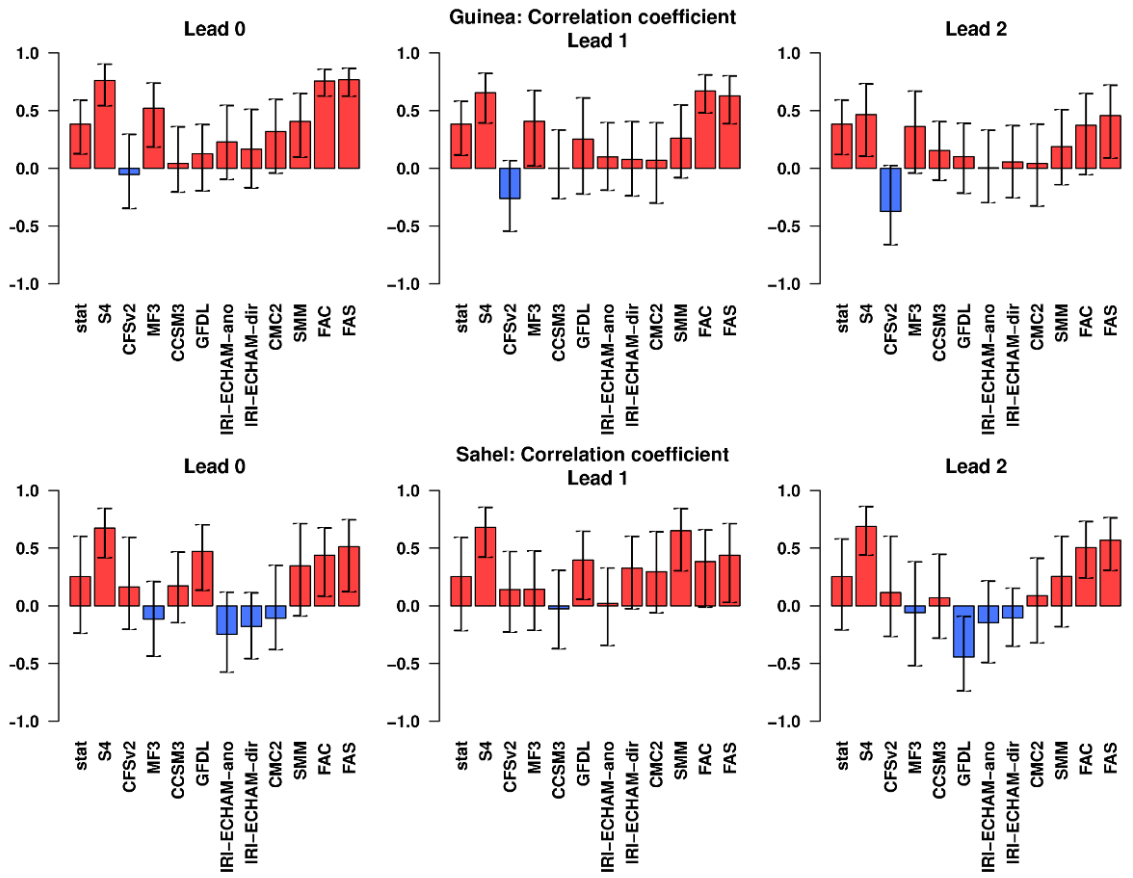

Figure 5.8: Correlation coefficient between the observed and predicted ensemble mean PCs for the period 1982-2011. The correlation was computed for the Sahelian (lower panel) and Guinean (upper panel) rainfall regimes and for lead times zero, one and two. The bars in each histogram represent the forecast systems. The lower and upper bound of the bootstrapped confidence interval is displayed as vertical bars.

Formulating skilful and reliable probabilistic predictions, which are the main requirements for decision making (Jolliffe and Stephenson, 2012; Doblas-Reyes et al., 2013a), is still an issue for most of the forecast systems analyzed here for the WAM rainfall regimes (Figure 5.7, 5.9 and 5.10). S4 has the best probabilistic prediction in terms of BSS (considering the events "rainfall regime above the median" or "above the upper quartile"; not shown), the CRPSS (Figure 5.9) and the ignorance skill score (Figure 5.10) more often than not. S4 is clearly an outlier as it is the only forecast system that has skill in terms of CRPSS and ignorance skill score. Another outlier is the CCSM3, which is the worst forecast system in almost all cases. Two reasons could explain this behavior in CCSM3 concerning the probabilistic scores: the small number of ensemble members (six

members), which makes its forecasts overconfident, and the low accuracy and large systematic error, as described above (Figure 5.7). As in the case of the correlation coefficient, the negative skill of most forecast systems makes the combinations to perform worse than S4 alone.

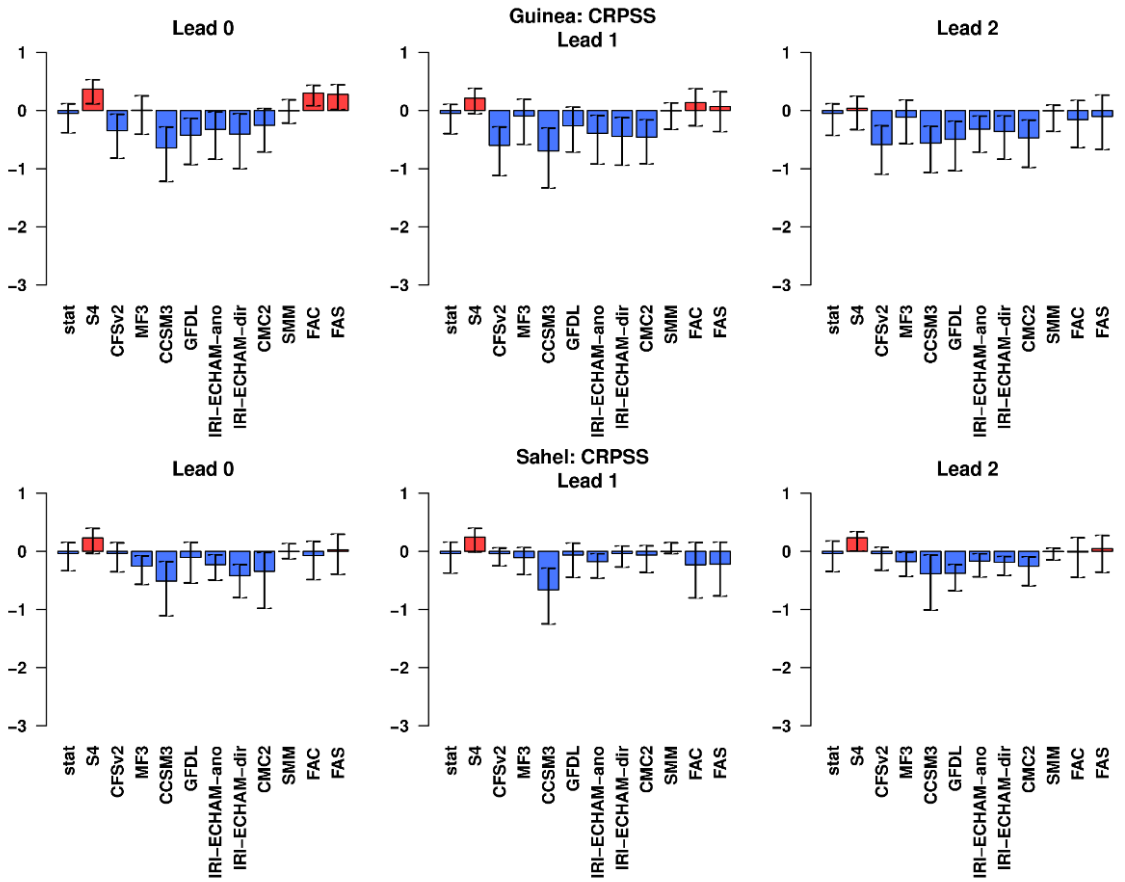

Figure 5.9: Same as Figure 5.8, but for the CRPSS.

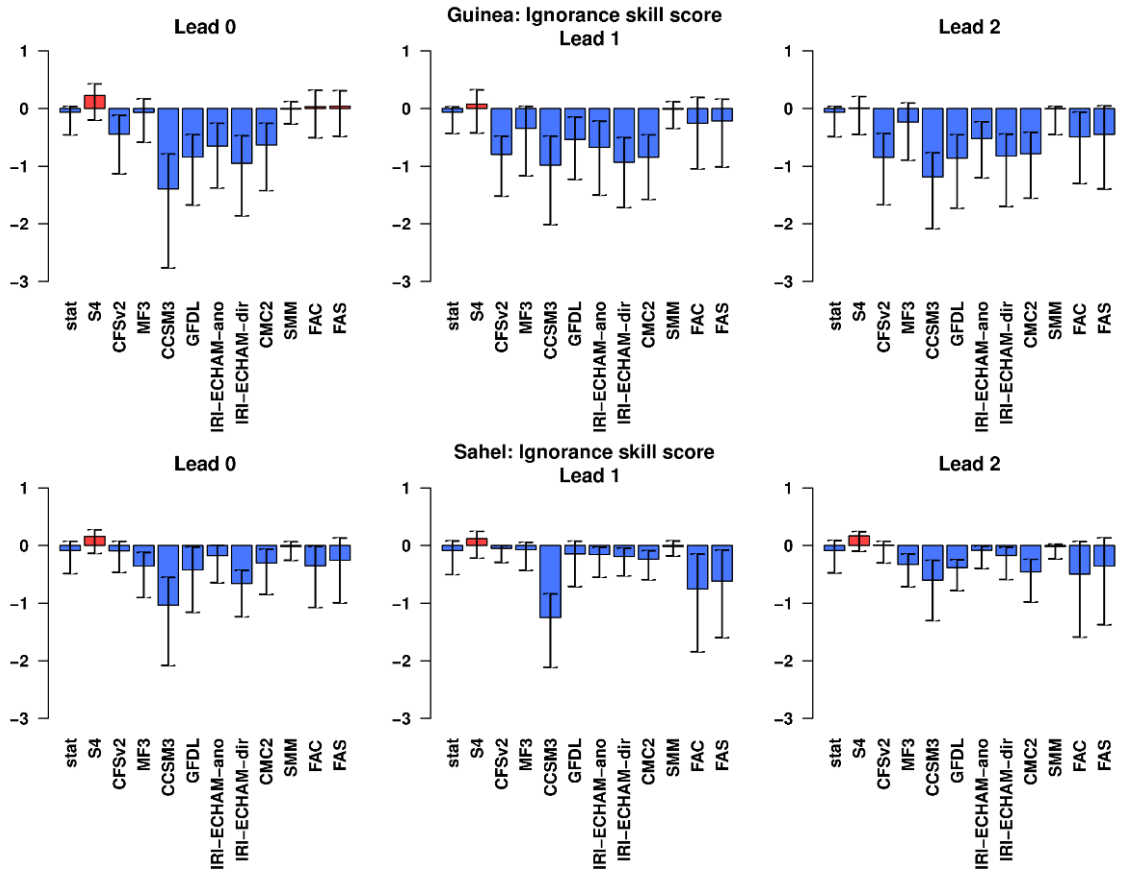

Figure 5.10: Same as Figure 5.8, but for the ignorance skill score.

5.3. Summary and conclusions

A targeted methodology to assess the year-to-year variations of the WAM rainfall variability has been illustrated in this chapter. This method estimates the main regimes of the WAM rainfall using monthly data averaged over 10ºW-10ºE covering the latitudes between the Equator and 20ºN and the period from June to October. The aim of the longitudinal averaging is to take into account the latitudinal migration and temporal distribution of the summer rainfall over the WAM region. This approach represents a process-oriented assessment of both the variability and predictability of the ITCZ-related WAM rainfall. PCA is applied on the seasonal evolution diagrams to estimate the leading modes of the WAM rainfall variability. PCA is performed upon the observations and each forecast system and lead time separately to take into account the fact that the hindcasts might represent the variability in a way different to the observations, while this representation also depends on the lead time (Doblas-Reyes et al., 2003; Philippon et al., 2010). The EOFs and the associated PCs related to the leading modes are used to describe the WAM rainfall regimes.

Two observational datasets (GPCP and GPCC) and a large number of quasi-operational forecast systems, among them the two systems from the EUROSIP initiative and six from the NMME project, are used in this chapter. The aim of using two different observational

datasets is twofold: firstly, to assess the observation uncertainty, and secondly, to build a statistical model using a dataset different from the one used for the forecast quality assessment. A simple statistical model built in retroactive mode as in an operational context (Mason and Baddour, 2008) is also used to predict the PCs associated with the Guinean and Sahelian regimes. Another aim of this research is to combine all the dynamical forecast systems and the statistical model to provide a single source of forecast information, something needed by the stakeholders (Doblas-Reyes et al., 2013a).

The forecast systems are combined using combination methods with both equal and unequal weights. In the first case, the predicted mean of each forecast system is averaged assigning equal weights to the forecast systems (i.e. simple average of the predicted mean). The second way of combining the forecast systems consists in assigning a larger weight to the systems that have smaller errors. The FA method (Coelho et al., 2004; Stephenson et al., 2005) is used to assign the weights. Finally, a forecast quality assessment is performed upon both combinations and forecast systems. Several deterministic and probabilistic verification scores have been used to take into account the high dimensionality of the forecast quality assessment (Murphy, 1991; Jolliffe and Stephenson, 2012). To the best of our knowledge, this work offers an unprecedented probabilistic evaluation of the seasonal prediction forecast quality of the WAM rainfall variability.

The main results of this chapter, which are innovative for the use of a large set of forecast systems and the way the seasonal variations of the WAM rainfall have been taken into account, are:

- As in previous studies (Motha et al., 1980; Fontaine et al., 1995; Fontaine and Janicot, 1996; Janicot et al., 1998; Giannini et al., 2003, 2005; Mohino et al., 2011b; Rodríguez-Fonseca et al., 2011), the two leading modes of the WAM rainfall variability are associated with the Guinean and Sahelian rainfall regimes. The Guinean and Sahelian regimes appear in the EOF1 and EOF2, respectively, when data are available over land and ocean (i.e. GPCP and the dynamical forecast systems). The Guinean (Sahelian) regime is found in the EOF2 (EOF1) when the data are available only over land (GPCP after applying a mask over the ocean and GPCC). For the common period 1982-2011, the variance explained by the Guinean mode varies from 29% (GPCP) to 20% (GPCC) and by the Sahelian mode from 31% (GPCP land-only) to 23% (GPCP) (Table 5.1).

- The PCs associated with the Guinean and Sahelian regimes estimated from GPCP are highly correlated with the ones estimated from GPCC. In addition, the PCs associated with the Guinean and Sahelian regimes estimated using a more traditional way, i.e. by applying a PCA on the spatial rainfall field, are highly correlated with the ones used in this study (Figure 5.5). This suggests that the seasonal variability does not modify the interannual nature of these regimes and that the substantial observational uncertainty is not as large as to substantially modify the characteristics of these regimes. The innovative component of the

analysis presented in this chapter is that the modes offer information about the intraseasonal variations of the rainfall regimes.

- Most forecast systems capture the main features associated with the Guinean regime (EOF1), that is, rainfall located south of 10ºN and the seasonal northward migration of rainfall. However, they are all biased and several of the forecast systems simulate the rainfall anomalies in the wrong location. On the other hand, only a fraction of the forecast systems capture the rainfall signal north of 10ºN associated with the Sahelian regime as observed in the GPCP dataset (EOF2).

- A fraction of the forecast systems have significant positive correlation (i.e., when the lower limit of the confidence interval is above zero) between the predicted mean and observed PC associated with the WAM regimes. However, only S4 has significant correlation when predicting both WAM regimes. MF3 performs well when predicting the Guinean regime and GFDL when predicting the Sahelian regime. The deterministic and probabilistic forecast quality assessment show two outliers: S4 and CCSM3. On the one hand, S4 is clearly the best forecast system for all scoring measures in most occasions. On the other hand, CCSM3 is clearly the worst system in most cases. Not surprisingly, it is shown that CCSM3 has the largest rainfall systematic errors over continental West Africa (Figure 5.3). CCSM3 has been identified as an outlier when compared to other NMME forecast systems in terms of root mean square error (RMSE) of tropical SST for September start dates and will be replaced by CCSM4 in the next phase of the NMME project (Kirtman et al., 2013).

- The simple statistical model outperforms several state-of-the-art dynamical forecast systems when predicting the PCs associated with the Guinean and Sahelian regimes (Figure 5.8). This result emphasizes the importance of using empirical benchmarks to compare with the dynamical forecast systems, particularly in an operational context.

- Combining all forecast systems do not lead to improved forecasts when compared to the best single forecast system, S4. In fact, S4 is far better than any forecast system when predicting the WAM rainfall regimes. This suggests that in some occasions, a multimodel approach is not necessarily better than an especially skillful model that is clearly identified.

Apart from showing that current operational or quasi-operational seasonal forecast systems can skillfully and reliably predict the interannual variations of the WAM rainfall regimes, which is an important result for the emerging climate services, the example described here illustrates that not always the SMM should be the preferred option in seasonal prediction. S4 is clearly the best forecast system when predicting both WAM rainfall regimes. The equal-weighting combination, with much lower skill than S4, does not improve the forecast quality of the resulting multimodel. At the same time, the two unequal-weighting combination approaches used here also do not improve the quality of the predictions with respect to S4. This suggests that the multimodel approach should not be automatically considered the best option in a prediction context and that a detailed analysis of the single systems should be carried out in each specific instance. Furthermore,

given the important investment in model and initial-condition development undertaken by ECMWF, it is clear that multimodel predictions will only improve if a sufficient number of single systems are continuously improved.

6. Prediction of near-surface temperature and precipitation over Europe

6.1. Introduction

Seasonal forecast skill has been traditionally low in the extratropical regions (Doblas-Reyes et al., 2013) because the response of the atmosphere to the slowly varying components of the climate system is difficult to be identified and modeled in these regions (Goddard et al., 2001; Scaife et al., 2014). Both statistical and dynamical forecast systems have been used to forecast surface climate variables at the seasonal time scale over extratropical regions (Goddard et al., 2001; Doblas-Reyes et al., 2013). Barnett and Preisendorfer (1987) studied the sources of forecast skill for monthly and seasonal-mean near-surface temperature over the United States using statistical models. They used sea level pressure (SLP), sea surface temperature (SST) and near-surface temperature itself as predictors in their statistical model. They found that forecast skill was generally low, except in January and February, and derived mainly from three sources: a decadal scale change in the northern hemisphere (NH) near-surface temperature, El Niño-Southern Oscillation (ENSO)-related phenomena and two non-ENSO short-lived, but large-scale structures in the pressure fields. Besides, persistence of the previously observed climate conditions prior to forecast issue time seemed to play an important role in summer (Barnett and Preisendorfer, 1987).

Using a similar methodology, Johansson et al. (1998) studied the seasonal forecast skill of near-surface temperature in northern Europe using 700 hPa geopotential height, SST and the predictand variable itself (lagged with respect to the predictand) as predictors. They showed that the highest skill is found in winter. For that season, geopotential height produces the highest skill due mainly to the North Atlantic Oscillation (NAO), but both near-surface temperature and local SST were also found to be important sources of skill. When quasi-global SST (i.e. 40ºS-60ºN, 180ºE-180ºW) was used as predictor much less skillful forecasts (in almost all target seasons) were obtained compared to the other predictors. However, when SST over six specific areas were used as predictors, they showed that the extratropical northern Pacific Ocean (30ºN-60ºN) SST produced the highest skill, pointing at the regional character of the SST influence. A secondary, weaker skill maximum was found in summer when only near-surface temperature was used as predictor (Johansson et al., 1998). Both studies showed that a combination of predictors in a multiple linear regression model did not improve forecast skill compared to simple linear regression models using a single predictor. Instead, the latter produced the highest skill when using ENSO as the single predictor for the North American climate (Barnett and Preisendorfer, 1987) and the NAO for the northern Europe climate (Johansson et al., 1998).

Blender et al. (2003) also attempted to build statistical models to predict monthly near-surface temperature, averaged spatially over three European regions (i.e. Great Britain, Germany and Scandinavia). From a variety of predictors, they chose two for each month and region that had the highest correlation with the predictand over the hindcast period to build a multiple linear regression model. The predictors tested were three teleconnection indices (i.e. the North Pacific pattern, the Southern Oscillation Index, and the NAO index), the first three principal components (PCs) of the North Atlantic SST, and monthly

anomalies of three surface climate variables (i.e. near-surface temperature, SLP and precipitation). They found that the skill has a seasonal cycle in the three regions, with maximum values observed in February-March and August-September. Besides, they showed that the near-surface temperature itself as predictor led to the highest forecast skill, with marginally positive correlation (up to 0.3) found in summer even at longer lead times (up to 11 months) in Great Britain and Scandinavia, but not in Germany.

Recently, Eden et al. (2015) assessed the forecast quality of several global linear regression models, built in retroactive mode, for seasonal near-surface temperature and precipitation with lead time one month. Global carbon dioxide equivalent concentration (CO_2Eq), Niño3.4, Pacific decadal oscillation (PDO), Atlantic Multidecadal Oscillation (AMO), quasi-biennial oscillation (QBO) and Indian Ocean Dipole (IOD) indices, as well as local SST and persistence of the predictand itself were all used as predictors. In contrast with the studies described above, in this case the multiple linear regression model generally outperforms the best simple linear regression model. However, when the CO_2Eq influence was removed, near-surface temperature skill was limited to certain regions, seasons and predictors. In this case, predictors such as persistence only led to skill over the oceans and continental areas in the tropics, while the Niño3.4 index led to skill over the ENSO teleconnection regions, such as tropical Pacific Ocean, northern Australia, Indonesia, Philippines, southeastern Africa, southern United States, northern and southeastern South America. For the extratropical North Atlantic/Europe sector, the AMO was the only predictor that led to near-surface temperature skill, while precipitation predictions were found to have skill limited to the ENSO teleconnection regions.

Using dynamical forecast systems, Doblas-Reyes et al. (2000) studied the seasonal forecast skill of the PRediction Of climate Variations On Seasonal to interannual Time-scales (PROVOST) multimodel ensemble over the NH. Three variables were analyzed: 850 hPa temperature, 500 hPa geopotential height and precipitation. The multimodel ensemble was made of three atmospheric general circulation models (AGCMs) integrated using prescribed SST, each of them having nine ensemble members. Several conclusions were drawn from that experiment. First, skill by single dynamical forecast systems varied with the variable and region. Similarly, the multimodel forecast skill in the midlatitudes showed a strong seasonality, with maximum values in winter. Second, forecast skill presents an interannual variation, with higher skill during ENSO years. Third, deterministic and probabilistic skill was spatially inhomogeneous and reduced over Europe when compared to other NH regions. Fourth, the multimodel forecasts presented higher probabilistic skill than individual models over both Europe and the NH. On the other hand, the multimodel ensemble mean did not always outperform individual forecast systems over Europe as it did over the NH. Some of these conclusions are similar to those described above using statistical models.

Similarly, Wang et al. (2009) applied the Climate Prediction and its Application to Society (CliPAS) multimodel ensemble, composed by 14 dynamical forecast systems, to study the global forecast skill of seasonal predictions of several global climate variables with lead time one month, both in winter and summer. Their conclusions endorsed the knowledge of the previous studies described above, such as the existence of a spatial distribution of the forecast skill with generally higher skill over the tropics than the

extratropics, including a limitation of forecast skill over Europe, as well as an interannual variability with virtually no skill in ENSO-neutral years. This later feature was also observed in the Development of a European Multimodel Ensemble Prediction System for Seasonal to Interannual Prediction (DEMETER) multimodel predictions of precipitation and maximum near-surface temperature over Spain (Frías et al., 2010). Using a different multimodel ensemble composed by three ACGMs, Mason et al. (1999) showed that the International Research Institute for Climate and Society (IRI) multimodel system has seasonal forecast skill when predicting precipitation and near-surface temperature for the strong 1997/98 El Niño event, except over Europe.

Given the importance of the coupling between the atmosphere with the slowly varying components of the climate system to seasonal prediction (Mason et al., 1999; Doblas-Reyes et al., 2000; Wang et al., 2009), Graham et al. (2005) compared the UK Met Office´s one-tier (GloSea3) and two-tier (HadAM3) seasonal forecast systems for probabilistic predictions of near-surface temperature in the upper tercile. The two-tier forecast system was integrated using persisted SST as boundary conditions. Four start dates (February, May, August and November), two lead times (one and three months), three regions (tropics, extratropics and Europe) and four scores (area under the relative operating characteristic (AROC), Brier Skill score (BSS), and its reliability and resolution components) were considered. They showed that GloSea3 systematically outperformed HadAM3 in all cases over the tropics. Over the extratropics, and more specifically over Europe, GloSea3 showed higher BSS than HadAM3 more often than not, especially due to improvements in the reliability term of the Brier score (BS). However, the BSS is negative over the extratropics in most cases (i.e. worse than the reference climatological forecast).

In a later study, the probabilistic forecast quality of the GloSea4 system to predict global SST, near-surface temperature and precipitation in winter and summer with lead time one month was assessed (Arribas et al., 2011). As in GloSea3, GloSea4 presents forecast skill over many tropical and subtropical regions, but not over extratropical regions. Arribas et al. (2011) noted that when the verification measure was the BSS, GloSea4 produces negative values for most land regions for both near-surface temperature and precipitation, which could be explained by the fact that the climatological forecast used as reference in the BSS is perfectly reliable (although having zero resolution) while GloSea4 was far from it. Therefore, the BSS will only be positive when a forecast system has good resolution, which makes the BSS a very demanding score for seasonal forecasting. Despite all the challenges in predicting the extratropical climate, Scaife et al. (2014) have recently shown that the GloSea5 has significant positive correlation skill when predicting the extratropical winter climate with lead time one month. They claimed that much of the skill found over North Atlantic and Europe is derived from the ability of this forecast system to predict the winter NAO, which in turn derived from an improved representation of the slowly varying components of the climate system compared to its predecessors. Despite the significant positive ensemble-mean correlation skill when predicting the NAO, the signal-to-noise-ratio (defined there as the standard deviation computed using the ensemble International Research Institute for Climate and Society mean divided by the standard deviation computed using all ensembles over the hindcast period) is low (Scaife et al., 2014). This means that the skill is sensitive to the ensemble size.

It has been shown that both statistical and dynamical forecast systems have limited forecast skill when predicting the surface climate over extratropical regions, especially over Europe. Even the combination of several forecast systems hardly led to any skill improvement. As the climatological prediction of precipitation and near-surface temperature is hard to beat in the extratropics, it makes sense to think of a way to combine predictions performed by complementary dynamical forecast systems with the climatology. In this respect, Robertson et al. (2004) applied a Bayesian methodology to combine several dynamical forecast systems with a climatological prior for tercile-category probabilities of seasonal-mean precipitation and near-surface temperature. They estimated weights for each forecast system, season, variable and land grid point independently having the ranked probability skill score (RPSS) as a skill measure. A spatial smother was applied to the weights of each forecast system with the aim of reducing the noise at the regional scale. The idea is that when no forecast system has skill in the RPSS sense at a certain season, variable and grid point, the final prediction would tend to the climatological forecast (e.g. for a variable well predicted by current forecast systems such as near-surface temperature, less weight is assigned to the climatological forecast and more weight to the better forecast systems). They concluded that this Bayesian combination scheme improves precipitation and near-surface temperature forecasts over the simple multimodel (SMM), which in turn, improves the forecasts over the single forecast systems. In the extratropics, the main benefit of this method is to bring much of the large area of negative precipitation RPSS values to near-zero values.

Using a different approach, Coelho et al. (2006) also applied a Bayesian methodology to combine several dynamical forecast systems with a statistical model to predict summer mean precipitation over South America. This technique is known as Forecast Assimilation (FA; Coelho et al., 2004, 2006; Stephenson et al., 2005). The FA technique applies a dimension reduction technique to deal with the high dimensionality of a multimodel ensemble of spatial fields with strong dependency between values of neighboring grid points (Stephenson et al., 2005), something not dealt with in Robertson et al. (2004), and that in a certain way substitutes the spatial smoothing of the weights. Coelho et al. (2006) showed that the combination of dynamical forecast systems with a statistical model, which have comparable precipitation skill level over South America, improved the skill over the SMM combination and the individual forecast systems in terms of BS and its components.

In order to contribute to the challenging task of predicting European climate as highlighted in previous studies described above, this chapter assesses the forecast quality of several statistical and dynamical forecast systems for near-surface temperature and precipitation predictions over Europe. Two start dates (i.e. May and November) and four lead times (i.e. zero through three months) are considered to represent monthly-anomalies in two seasons: summer (MJJA) and winter (NDJF). Statistical models based on simple linear regression are also estimated using four different predictors, the predictand itself lagged in time and three SST indices (Niño3.4, subtropical North Atlantic (SNA) and AMO). The dynamical forecast systems are combined using two different combination methods: the SMM and the FA method. The benefits and limitations of combining the

forecast systems using these two methods are discussed. An overall assessment of the skill of the combinations and the single forecast systems is shown.

6.2. Forecast quality assessment

In this section, the forecast quality assessment of the near-surface temperature and precipitation predictions in summer and winter by multiple forecast systems will be discussed.

6.2.1. Statistical model

Statistical models based on simple linear regression as described in Coelho et al. (2004) are estimated using four different predictors: the local predictand itself and three SST indices (Niño3.4, SNA and AMO). These predictors are used to predict monthly near-surface temperature and precipitation in summer (MJJA) and winter (NDJF). A combination of predictors has not been taken into account because it had been shown in previous studies that this does not improve forecast skill (Barnett and Preisendorfer, 1987; Johansson et al., 1998), except when CO_2Eq is also used as predictor (Eden et al., 2015). For a fair comparison with the dynamical forecast systems, the predictors of the month of April are used to predict near-surface temperature and precipitation in May, June, July and August, and the predictors of the month of October to predict these variables in November, December, January and February. Thus, the four lead times considered are zero to three months.

The statistical models are built in retroactive mode, that is, only years prior to the target period are used in the estimation of the regression coefficients as in an operational context (Mason and Mimmack, 2002; Mason and Baddour, 2008; Eden et al., 2015). The first training period 1951-1981 is increased by one year at a time to predict the target years from 1982 to 2010 (the same period available for all hindcasts of the dynamical forecast systems). In this study, GHCNv2 (near-surface temperature) and GPCC (precipitation) datasets are used to train the statistical models over continental areas to take advantage of their long time series to train the statistical models in retroactive mode. However, the forecast quality of both statistical and dynamical forecast systems is assessed using a common data, that is, GPCP for precipitation and ERA-Interim for near-surface temperature. The SST indices used as predictors in the statistical models are computed using ERSST.

As shown in previous studies (e.g. Barston, 1994; Lang et al., 2014), forecast skill is more sensitive to the predictand, predictor, target region and season than to lead time. That is, forecast skill varies little from lead time zero to four (Appendix C). Another well-known result is that regardless of the predictor, the statistical models predict better near-surface temperature than precipitation (Figure 6.1). In fact, none of the four predictors used to predict precipitation leads to statistically significant correlation, except for small areas. As noted in previous studies, predicting precipitation outside the ENSO teleconnection areas using linear models is a very difficult goal (e.g. Eden et al., 2015). This is because linear models do not account for the non-linear interactions between the atmosphere

region and the slowly varying components of the climate system in the extratropical region, which limits their usefulness.

Near-surface temperature predictions in summer are more skillful than in winter (two right columns of Figure 6.1). Much of this skill in summer months might be derived from the near-surface temperature trend between 1982 and 2010, something not detected in winter (EEA, 2015). The slope of near-surface temperature linear trend has been computed for each grid-point and target month independently using ERA-Interim dataset (Appendix D). We found a statistically significant warming trend in May and June over the Western Europe, the Mediterranean Sea and northern Africa. On the other hand, during the months of July and August warming trend is detected in northern and eastern Europe, northern Africa, Eurasia, and the Middle East. These are the areas where the statistical models yield statistically significant correlation skill in July and August (Appendix C). The warming trend in the summer months might also explain the similarity among the correlation patterns of the linear regression models built with different predictors (third column of Figure 6.1 and second column of the first row of Figure D1). No warming or a cooling trend is observed over continental Europe in some winter months (second row of Figure D1), which might be linked to the positive snow cover trend over Eurasian region (Cohen, 2011). Eden et al. (2015) showed that the forecast skill is considerably reduced when the trend is removed before estimating the regression coefficients in linear regression models.

6.2.2. Dynamical forecast systems

As with the statistical models, dynamical forecast systems are usually more skillful when predicting near-surface temperature (Figure 6.2) than precipitation (Figure 6.3). Figure 6.2 shows the correlation coefficient between predicted and observed near-surface temperature in June, with the predictions initialized in May (lead time 1) for the hindcast period 1982-2010. Four out of nine forecast systems present statistically significant positive correlation over the North Atlantic, continental Europe and the Mediterranean Sea. These forecast systems are S4, CFSv2, GFDL, and NASA. MF3 and CMC2 display statistically significant positive correlation in the northern North Atlantic, southern Europe and the Mediterranean Sea, whereas CCSM3 and the IRI-ECHAM forecast systems display negative or non-significant positive correlation in most areas over continental Europe. On the other hand, precipitation forecast skill in December for predictions initialized in November display a different picture (Figure 6.3): only a small fraction of the forecast systems has positive correlation and, among these ones, statistically significant positive correlation is located mostly over the ocean. As explained in the introductory section, low ow skill of dynamical forecast system when predicting surface climate variables in the extratropics has been shown in previous studies (Doblas-Reyes et al., 2000; Wang et al., 2009; Arribas et al., 2011).

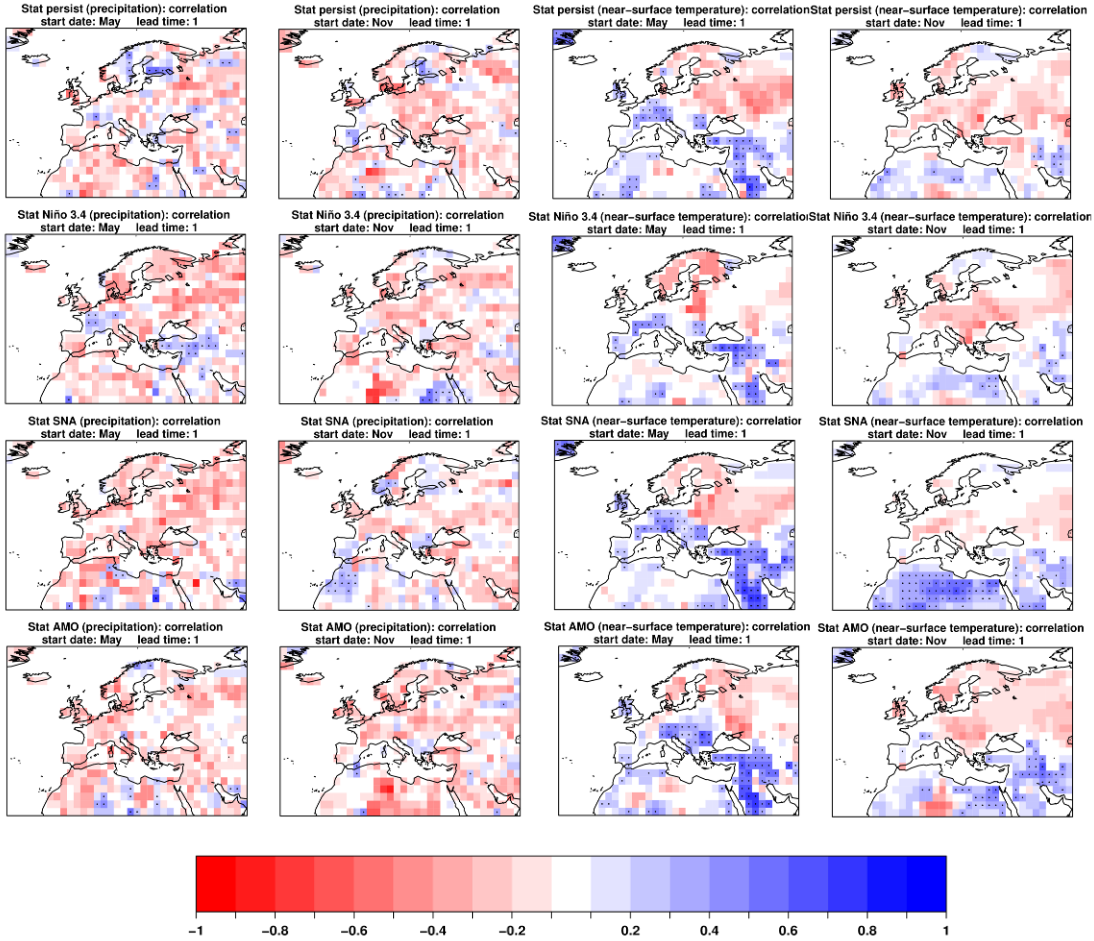

Figure 6.1: Correlation between predicted and observed precipitation (first and second columns) and near-surface temperature (third and fourth columns) in June (first and third rows) and December (second and fourth rows). The correlation was computed for the hindcast period 1982-2010. The statistical model was estimated using four different predictors: the predictand variable itself at the same grid point, SST Niño3.4, SNA and AMO indices. Anomaly values of April and October are used to predict June and December, respectively. The regression coefficients were estimated in retroactive mode. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

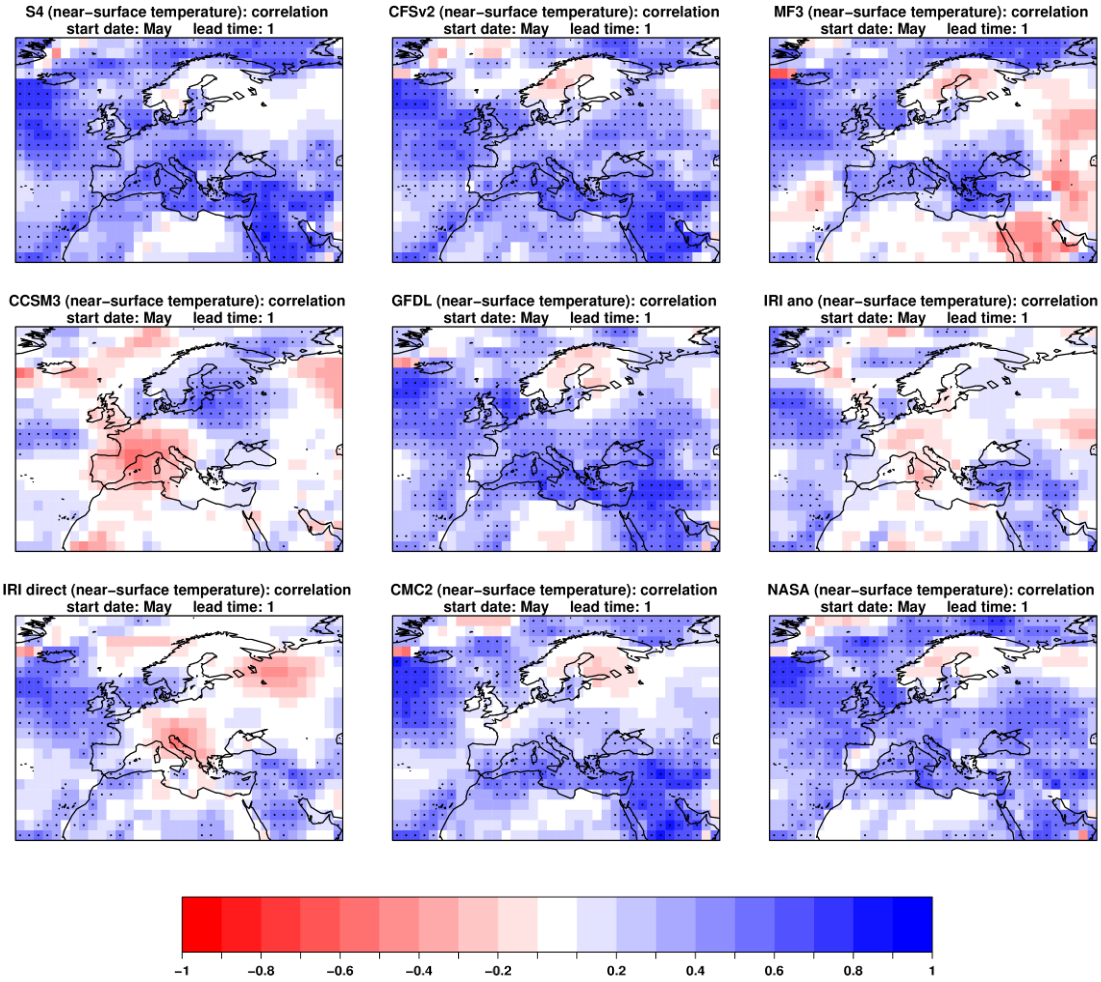

Figure 6.2: Correlation coefficient between predicted and observed near-surface temperature in June. Predictions are for May start dates (lead time 1). The correlation was computed for the hindcast period 1982-2010. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

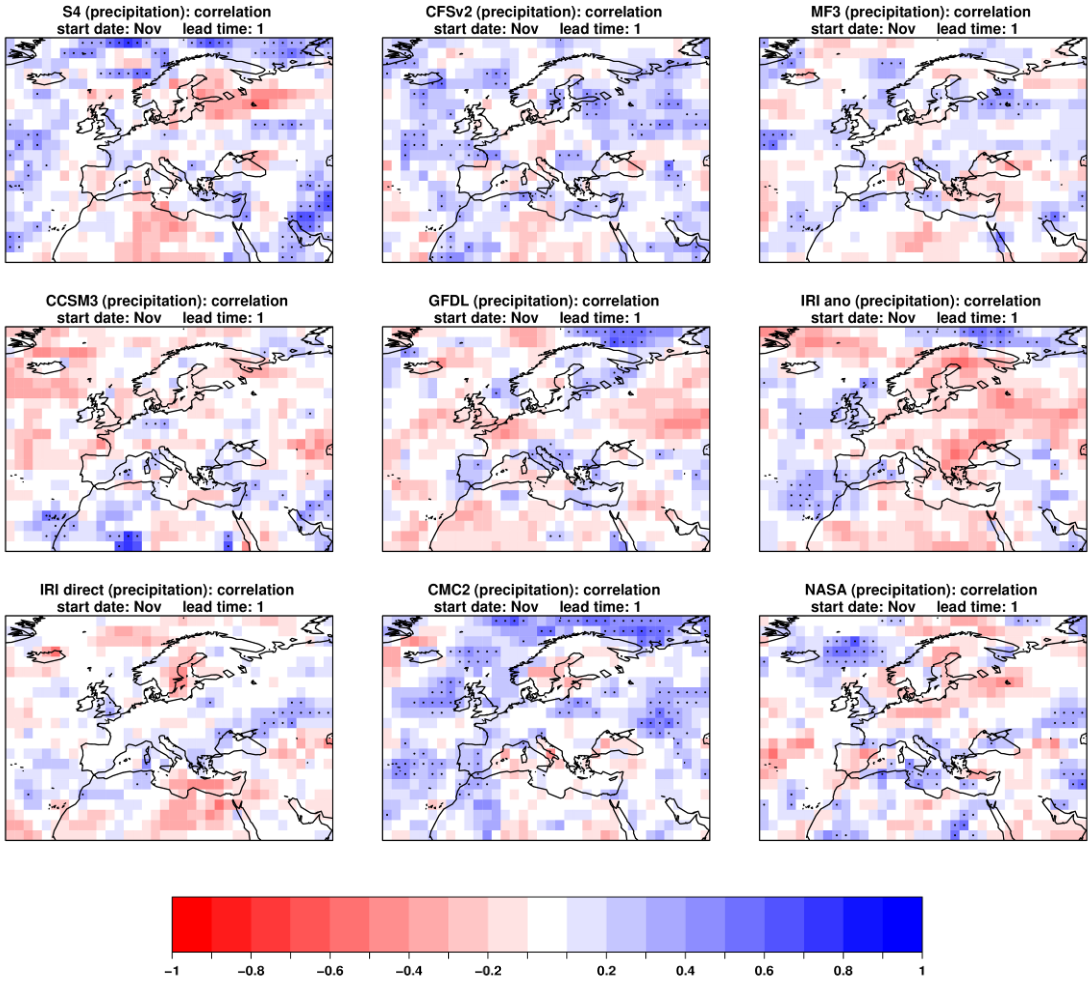

Figure 6.3: Correlation coefficient between predicted and observed precipitation in December. Predictions are for November start date (lead time 1). The correlation was computed for the hindcast period 1982-2010. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

The precipitation skill decreases considerably with lead time. Figure 6.4 illustrates the correlation coefficient between the S4 ensemble mean and observed precipitation decreasing rapidly from lead time 0 to lead time three months, both in summer and winter. S4 is one of the forecast systems, together with CFSv2, CMC2 and NASA, that has statistically significant positive correlation coefficient in a large number of grid points at lead time 0 (not shown). Some of the forecast system do not display statistically significant precipitation skill even at lead time 0. The decrease in correlation with increasing lead time is also observed for near-surface temperature predictions in winter, but not in summer. This is probably due to the trend in the observed near-surface temperature in summer months (Appendix D), a feature captured by many of the forecast systems, including S4.

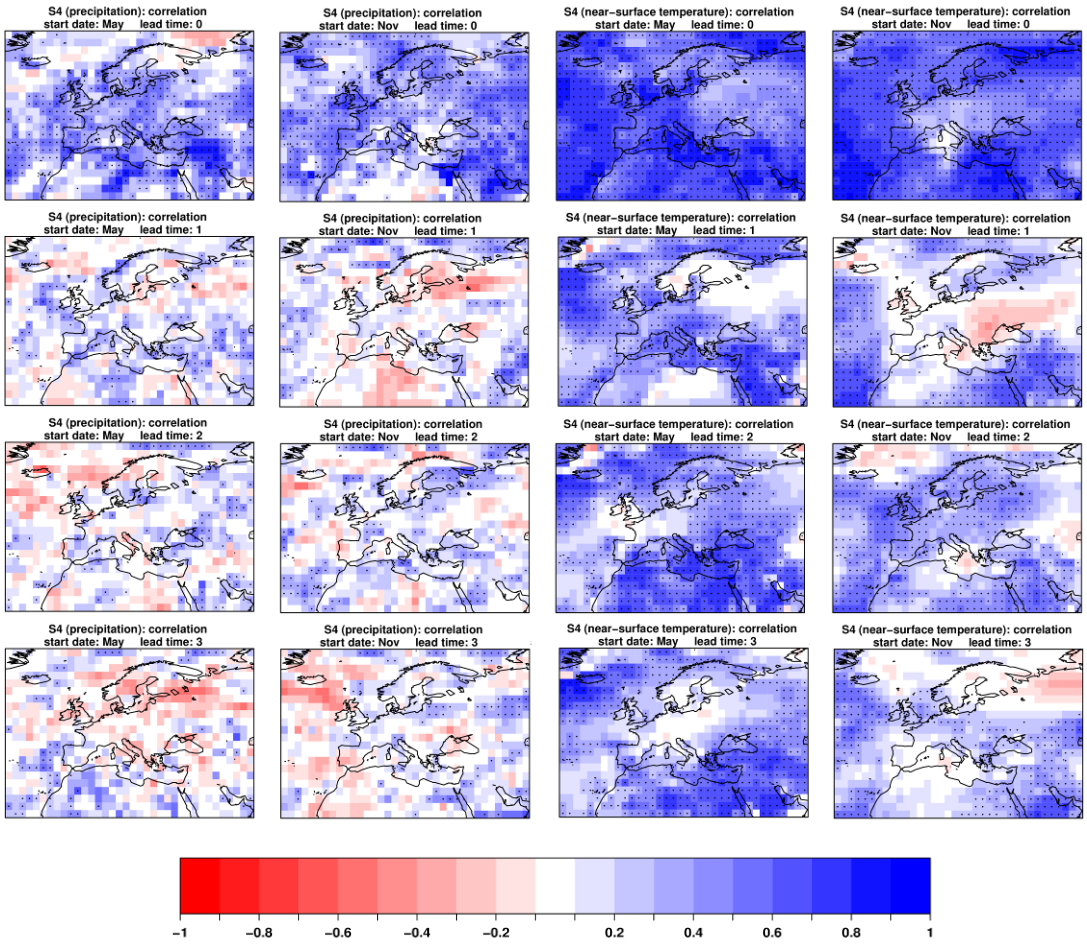

Figure 6.4: Correlation coefficient between predicted and observed near-surface temperature (two right columns) and precipitation (two left columns) in May, June, July and August for predictions initialized in May (first and third column) and in November, December, January and February for predictions initialized in November (second and fourth column). The correlation was computed for the hindcast period 1982-2010. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

Due to the high dimensionality of the forecast verification problem, it is very important to take into account multiple verification measures to obtain richer and more robust conclusions about the quality and/or value of the forecast systems (Murphy, 1991; Mason and Stephenson, 2008). Therefore, we have also used one probabilistic score to verify how well forecast uncertainty (e.g. the spread of an ensemble system) is predicted by current forecast systems. The CRPS is used to measure the quality of the probabilistic predictions and serves as complementary information to the correlation coefficient. The CRPS is displayed as skill scores where the climatology is the reference forecast. In this case, the climatological CDF is built up by considering each hindcast year, except the target year and one year prior to and one after it, as ensemble members. The CRPSS is positive when the forecast system has better (smaller) CRPS than the climatological naïve forecast.

The CRPSS estimated for near-surface temperature predictions in June at lead time one month (Figure 6.5) illustrates how hard it is to beat the climatological forecast (Feddersen and Andersen, 2005; Graham et al., 2005). None of the forecast systems displays significant positive CRPSS over continental Europe and only S4 and CFSv2 display more positive than negative CRPSS areas over the studied region. Even at lead time 0 (May), only three out of nine forecast systems (S4, CFSv2 and NASA) have more positive than negative CRPSS grid points. A similar picture is found in July and August as well as in winter months (not shown). The CRPSS for precipitation predictions shows an even more pessimistic view of the forecast quality of current forecast systems (not shown). Only S4 and CFSv2 present positive CRPSS at lead time 0, both in summer and winter. However, all the skill goes to zero or negative at lead times one, two and three months. It should be borne in mind that the CRPSS is a more stringent skill measure than the correlation because it requires not only that the signal has the correct sign (as the correlation is not sensitive to errors in the forecast variability), but also that the uncertainty is correctly predicted, which is usually not the case in current systems.

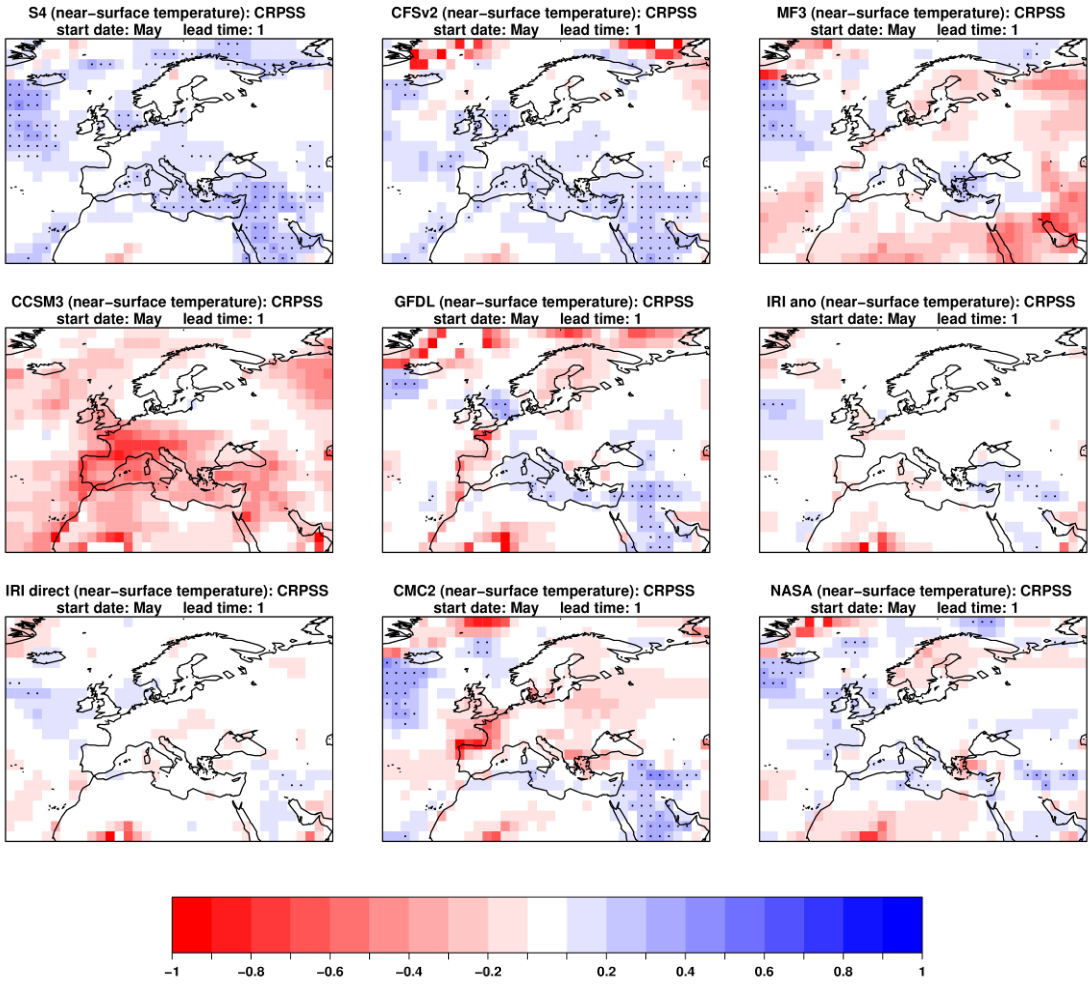

Figure 6.5: CRPSS for near-surface temperature predictions in June. Predictions are for May start dates (lead time 1). The CRPSS was computed for the hindcast period 1982-2010. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

The ignorance score, which is defined as the negative logarithm of the predicted probability density of the observation, was also computed to quantify the quality of the predicted uncertainty (Roulston and Smith, 2002). A kernel density estimation using the ensemble members for each individual forecast system is used to quantify the predicted uncertainty. As the CRPS, an ensemble forecast system gets a perfect ignorance score when all ensemble members equal the observation (i.e. the predicted probability density equals one and its logarithm equals zero). On the other hand, the logarithm goes to infinity when zero probability density is assigned to a value that was actually observed. This penalty, which is often applied to unreliable ensemble forecast systems with a small number of members, is more severe than in the CRPS. As with the CRPSS, only S4 and CFSv2 predictions get positive ignorance skill score for near-surface temperature predictions at lead time 0, which rapidly decreases to near zero or negative values at leads longer than one month (not shown). The other forecast systems get near zero or negative ignorance skill score in most cases (i.e. month being predicted, lead time and grid point). For precipitation predictions, most forecast systems are strongly penalized for assigning large probabilities to values that do not occur, which leads to large negative values in the ignorance skill score. The ignorance skill score usually shows a more pessimist of the forecast quality when compared to the CRPSS; thus, only the latter score has been illustrated in this chapter.

6.3. FA predictions

6.3.1. Modes of variability

As described in Chapter 3, the first step to perform the combination of several dynamical forecast systems using the FA technique is to apply the MCA on the gridded data (Stephenson et al., 2005; Coelho et al., 2006). The MCA is applied on the cross-covariance matrix of the observations and predictions, where the left matrix has the longitudes and latitudes of the observations on the rows, and the right matrix has the longitudes and latitudes of the ensemble mean of all forecast systems on the rows. Both matrices have the target years on the columns. This is necessary because the number of grid points is much higher than the number of independent forecasts and the dependency between values at neighboring grid points (Stephenson et al., 2005). Therefore, the combination is first performed in the MCA space and then brought back to the longitude-latitude space. In this section, the modes of variability estimated from the MCA analysis applied to the cross-covariance matrix between the forecasts and observations, used to estimate the parameters needed to combine several forecasts using the FA method, will be described from a physical point of view. The MCA analyses were performed independently on two variables (i.e. precipitation and near-surface temperature), two start dates (i.e. May and November), four lead times (i.e. zero to three months), and nine dynamical forecast systems. The large amount of cases makes a detailed description of every mode of variability unfeasible. Therefore, only one example will be shown.

The correlation between the expansion coefficients (time series) associated with the MCA variability modes and several climate indices is computed to quantify the relation between the MCA modes and the main teleconnection patterns of the European climate variability. Some of the teleconnection patterns that have influence over the European climate are analyzed, including the NAO, Artic Oscillation (AO), East Atlantic (EA), East

Atlantic/Western Russia (EAWR), Scandinavian and Polar/Eurasian (PE). The MCA variability modes will be illustrated below as the heterogeneous correlation maps computed by correlating the expansion coefficients of the left field (i.e. the observation) with the original data of the right field (i.e. the forecast systems) and vice versa. The aim is to compare several variability modes with the same standard of units.

The first observed mode of near-surface temperature variability in June displays positive values over most of the area, except for a few grid points in the Eurasian region and in the Norwegian Sea (top left panel of Figure 6.6). The expansion coefficient associated with this mode (red line of the left panel of Figure 8) has null or near null correlation with physical teleconnection patterns described above. S4, CFSv2, GFDL, CMC2 and NASA (first column of Figure 6.6) simulate this widespread positive near-surface temperature, but they overestimate the magnitude of the values compared to observations. The IRI-ECHAM systems predict a generalized temperature increase, but shift westward the negative values over Eurasia compared to observations. This variability mode accounts for 51% of the squared covariance between the observations and the predictions at lead time one month.

The second observed near-surface temperature variability mode, which accounts for 19% of the squared covariance, displays a strong dipole with positive anomalies over central Europe and negative anomalies centered over Russia (top central panel of Figure 6.6). The expansion coefficient associated with this variability mode (red line of the central panel of Figure 6.10) has a small correlation with the NAO (0.31), AO (0.34) and EA (0.40), and near null correlation with the other analyzed teleconnection indices. These three climate indices are then correlated with the observed grid-point near-surface temperature anomalies for the period 1982-2010. The correlation between EA and the near-surface temperature anomalies exhibit a pattern similar to the second variability mode displayed in the top central panel of Figure 6.6 (not shown). All forecast systems underestimate the magnitude of the anomalies associated with this variability mode, and most of them fail to simulate it (second column of Figure 6.6). As an exception to this, S4 simulates a similar dipole, but the positive anomalies over central Europe are shifted southward whereas the negative anomalies over Russia are shifted northward. In addition, S4 captures a similar pattern compared to observation over the North Atlantic Ocean, central Africa and the Middle East. All other forecast systems failed to simulate the negative anomalies over Russia, except for IRI-ECHAM anomaly that simulates anomalies of the opposite sign. Similarly, most forecast systems, with the exception of GFDL and CMC2 yet with considerably biases, have difficulties in simulating the positive anomalies over continental Europe and the Mediterranean Sea.

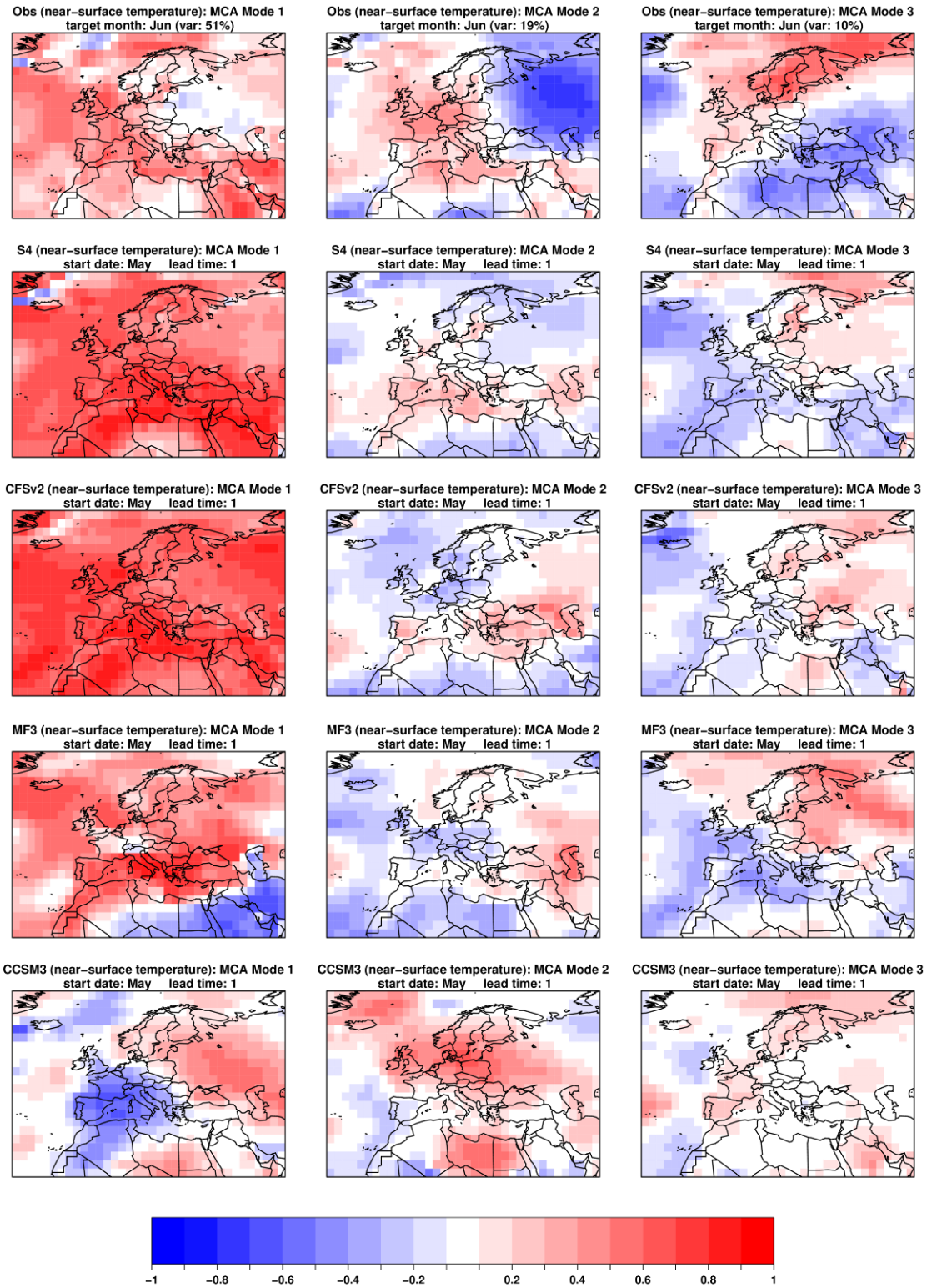

Figure 6.6: Observed and predicted heterogeneous correlation maps of near-surface temperature in June. Predictions are for May start date (lead time 1). The expansion coefficients of the left field (i.e. the observation) are correlated with the original data of the right field (i.e. the forecast systems) and vice versa. Results are shown for the first three leading MCA modes, computed for the hindcast period 1982-2010.

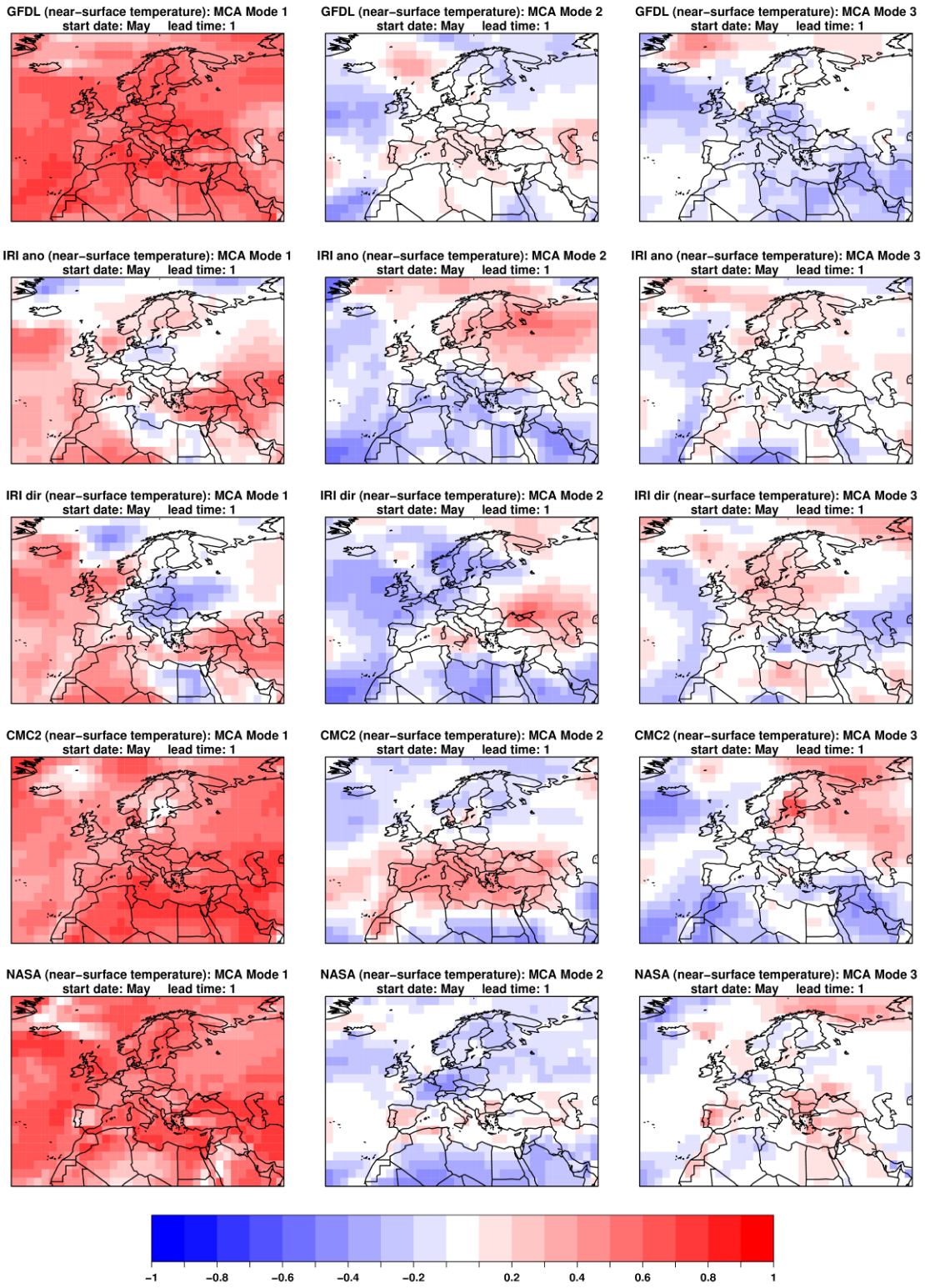

Figure 6.6: Continue.

The third observed variability mode (top right panel of Figure 6.6) shows a pattern with positive anomalies over most of Europe and the Norwegian Sea and negative anomalies over some parts of North Atlantic Ocean, northern Africa, and Middle East. This mode accounts for 10% of the squared covariance. None of the climate indices described above

has correlation higher than 0.3 with the expansion coefficient associated with this variability mode, which makes it difficult to associate it with a main teleconnection pattern. All forecast systems capture the negative anomalies over the North Atlantic Ocean, but most of them fail to predict the pattern over the continents. The only forecast system that predicts a variability that slightly resembles the observed one over the continents is the IRI-ECHAM direct.

The first three observed MCA modes of near-surface temperature variability in May, July and August have similar patterns to the ones in June (first row of Figure 7). However, the variance explained by each mode varies considerably from May (predictions at lead time zero) to August (predictions at lead time three months): while the first three modes account for 33%, 24% and 17% of the squared covariance in May, these numbers change to 66%, 10% and 7%, respectively, in August. That is, not only there is an increase in the squared covariance accounted for these three modes with lead time, but also the share of the covariance accounted for the first mode alone doubles while the share of the covariance accounted for the second and third modes are reduced by half each. If the observed variability patterns do not vary with lead time, then the explanation for the changes in the accounted covariance might come from the predictions. In fact, most of the forecast systems consistently increase the positive anomalies of the first variability mode with lead time which could be associated with the different trends simulated for different lead times.

The first observed mode of near-surface temperature variability in December shows widespread positive anomalies over most Europe, Eurasia and Middle East, and negative anomalies over the Iberian Peninsula and Iceland (first column of the first row of Figure E1 of the Appendix E). This pattern is different to the observed near-surface temperature trend in December where a cooling trend is observed over most of Europe, not limited to the Iberian Peninsula (second column of the second row of Figure D1 of the Appendix D). This difference might be explained because some of the forecast systems used in the cross-covariance matrix of the observations and predictions used to estimate the variability modes have different trend at this lead time. The expansion coefficient associated with this variability mode has correlation with the NAO (0.43), AO (0.38), AE (0.30) and Scandinavian (-0.30) indices. The correlation between the grid-point near-surface temperature anomalies and the NAO/AO both in December have many similar features to the variability mode shown in top left panel of Figure E1. These features are consistent with the weakening of the meridional circulation associated with the negative phase of NAO/AO. On the other hand, the EA pattern presents positive anomalies over most of Western Europe and the Scandinavian pattern presents negative anomalies over the Iberian Peninsula and positive anomalies over eastern Mediterranean. Neither the pattern nor its explained variance change considerably with lead time (it changes from 69% to 63%, from lead time zero to three months).

The second variability mode in November, with variance 10%, displays negative anomalies in northern Europe and norther Russia, and positive anomalies over North Atlantic, Southern Europe and Middle East. The expansion coefficient associated with this mode is also positively correlated with the NAO (0.36), which pattern in November resembles that of the top central panel of Figure E1. Most forecast systems fail to capture

this variability mode, with the exception of CCSM3 that displays some of the features present in the observed mode (Figure E1). This failure might be explained by the difficulties of these forecast systems to predict near-surface temperature in December with lead time 1 month (Figure 6.2) whose forecast skill over Europe is even worse than at lead times two and three months (Figure 6.4). The third variability mode has positive anomalies over central and northern Europe, the eastern side of the Mediterranean Sea and northeast Africa and negative anomalies over the Iberian Peninsula and northwestern Africa (top right panel of Figure E1). The expansion coefficient associated with this variability mode, which explains only 7% of the total variance, has positive correlation with the EA index (0.41). However, the spatial EA pattern in December is very different from the one in the top right panel of Figure E1. Forecast systems fail to predict some of the features found in the observations, especially over Europe and western Mediterranean. Differently from the summer months, the variances explained by these three modes do not change significantly with lead time.

Figure 6.7 illustrates the first three observed and predicted precipitation variability modes for May, which explains 18% of the total variance. Predictions are for lead time 0 (i.e. produced in early May and valid for May). The first precipitation variability mode from observations (top left panel of Figure 6.7) displays a wave-like pattern with a strong positive area centered over west Iceland, a strong negative area centered over Northern Europe, a weak positive area over Southern Europe, the Mediterranean Sea and Northern Africa, and a weak negative one over Northwestern Africa. The expansion coefficient associated with this mode has correlation with the NAO (0.37), AO (0.36) and EAWR (0.47) indices. The precipitation pattern associated with the EAWR index in May presents many similar features to the first observed precipitation variability mode. Some of the forecast systems capture reasonably well this pattern, such as S4 and CFSv2 (first column of Figure 6.7). Some others, such as MF3, GFDL, CMC2 and NASA forecast systems, reproduce the dipole Iceland-Northern Europe, but not the Southern Europe-Western Africa one. Although IRI-ECHAM anomaly reproduces the wave-like pattern, the pattern does not match that of the observations. IRI-ECHAM direct and CCSM3 do not capture the observed pattern.

The second observed variability mode (top central panel of Figure 6.7) also displays a wave like pattern with negative anomalies over the UK, North Sea and Norway, positive anomalies over from the North Atlantic Ocean off the coast of Africa to the Eurasian region, and another negative region with negative anomalies over Kazakhstan and Russia. This mode explains 13% of the total variance of the cross-covariance matrix of the observed and predicted precipitation. The expansion coefficient associated with this mode has negative correlation with the AO (-0.34), EA (-0.77) and EAWR (-0.40) indices. This negative correlation with the AO and EAWR indices is expected given the similarities between the wave patterns in the first and second variability mode. On the other hand, precipitation pattern associated with the AE in May is very similar to the one shown in the top central panel of Figure 6.7. Some of the dynamical forecast systems, such as S4, CFSv2, MF3, CMC2 and NASA, capture reasonably well this pattern (second column of Figure 6.7). IRI-ECHAM anomaly captures well the negative anomalies over Kazakhstan, but not the positive ones over Eastern Europe and the eastern bound of the Mediterranean Sea. CCSM3 and IRI-ECHAM direct predictions do not display a wave like pattern in this variability mode.

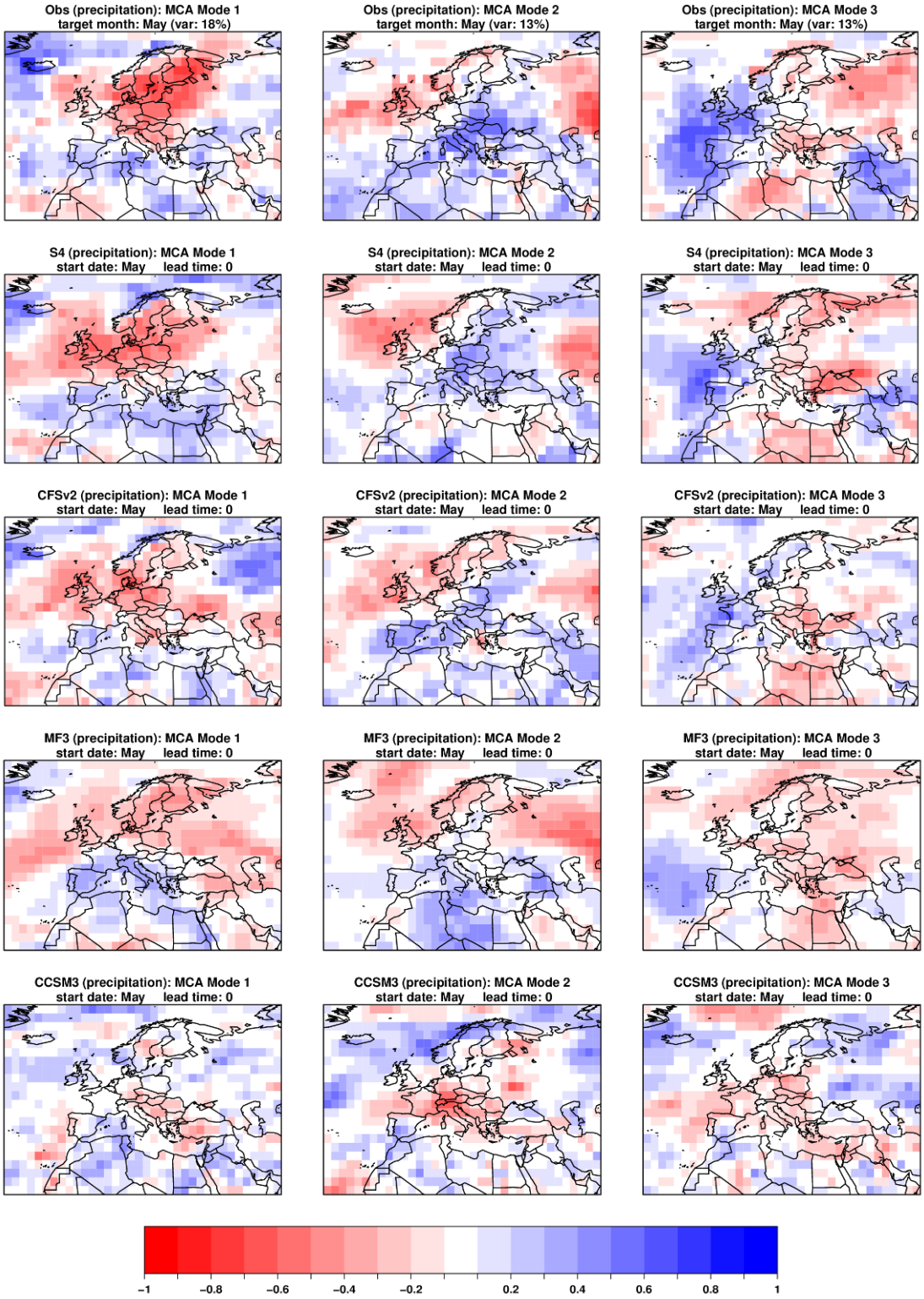

Figure 6.7: Observed and predicted heterogeneous correlation maps of precipitation in May. Predictions are for May start date (lead time 0). The expansion coefficients of the left field (i.e. the observation) are correlated with the original data of the right field (i.e. the forecast systems) and vice versa. Results are shown for the first three leading MCA modes, computed for the hindcast period 1982-2010.

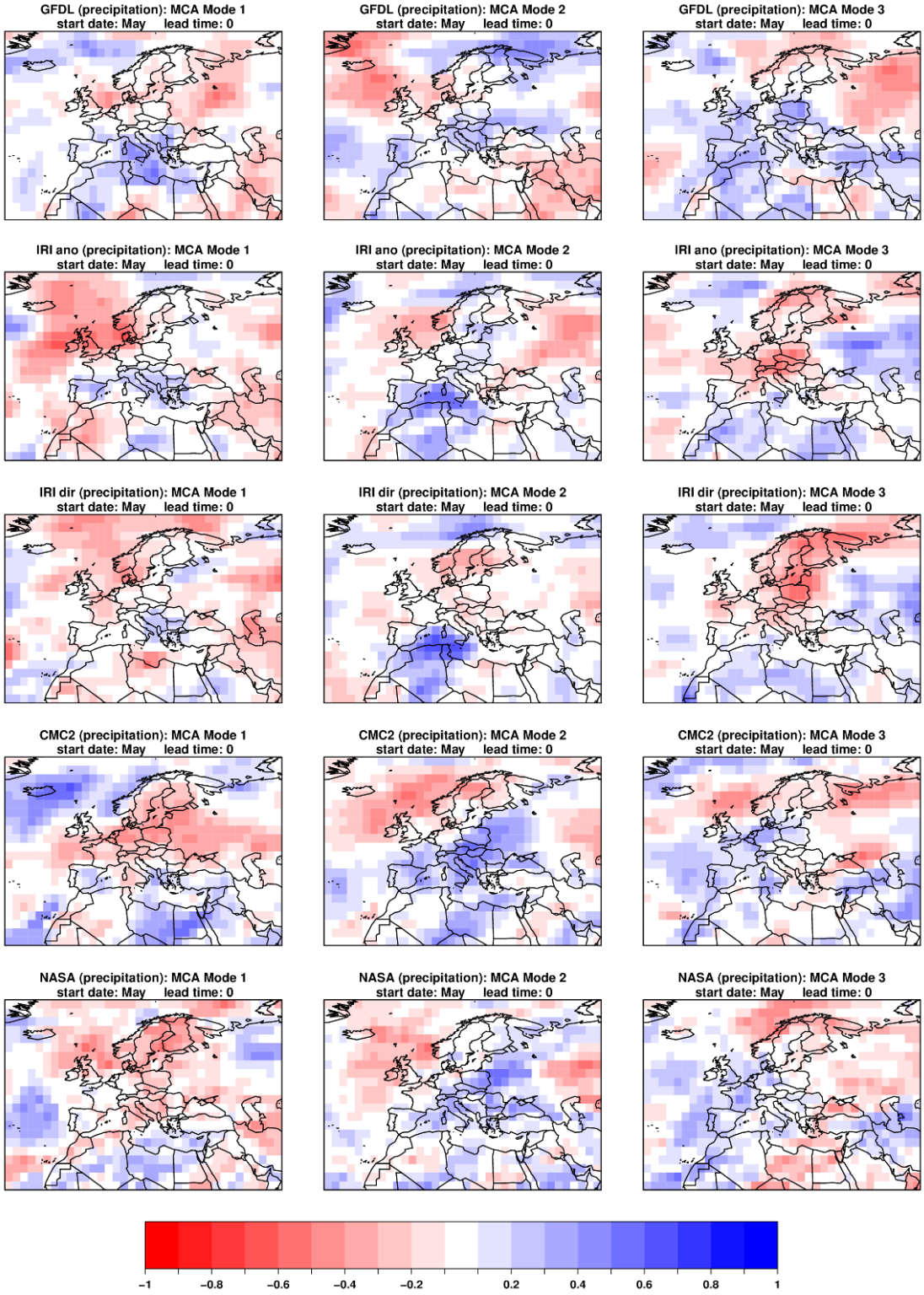

Figure 6.7: Continue.

The third observed variability mode (top right panel of Figure 6.7) displays another wave-like pattern, zonally oriented with positive anomalies over Western Europe and Northwestern Africa, a negative band that extends from Northern Africa to Russia and once more positive anomalies over the Middle East. The variance associated with this variability mode is 13%, equal to the variance explained by the second mode. The

expansion coefficient associated with this mode has negative correlation with the AO (-0.39) and positive correlation with the Scandinavian index (0.49). S4 is the forecast system that best simulates this mode of variability. CFSv2, MF3, IRI-ECHAM systems, CMC2 and NASA also simulate the observed wave-like pattern, but the magnitude and location of the anomalies differs from the observations.

The first three observed modes of precipitation variability in June (predictions at lead time one month), July (predictions at lead time two months) and August (predictions at lead time three months) display pattern with many similar features to the ones in May. This can be illustrated by comparing the first precipitation variability mode in July (top left panel of Figure 6.8) and its counterpart in May (top let panel of Figure 6.7). However, the expansion coefficient associated with the former is only weakly correlated with the NAO (0.01) and EAWR (-0.09) indices. This might be explained because most forecast systems at lead time two months fail to reproduce the dipole between negative (positive) precipitation anomalies in northern (southern) Europe resulting from a negative NAO and a dry (wet) pattern over Europe (North Atlantic) associated with a positive EAWR (left column of Figure 6.8).

The wave-like pattern of the second observed precipitation variability mode in July is hardly reproduced by the forecast system with predictions initialized two months in advance (central column of Figure 6.8). This mode explains 12% of the total variance. However, some of them capture at least in part of the features (e.g. wave-like pattern) found in the observed variability mode, although with significant differences in the position and magnitude of the anomalies. A comparison of the second column of Figures 6.7 and 6.8 illustrates how fast predicted variability mode drifts from the observations from lead time zero to lead time two. The shift in the position and magnitude of the anomalies is at the origin of low skill because one tends to estimate the skill locally and not averaged over large areas. For example, the second precipitation variability mode is well predicted in May with lead time zero by S4 (central column of the second row of Figure 6.7). The same forecast system predict a similar wave-like pattern in July, but the anomalies are shifted northward (central column of the second row of Figure 6.7). This shift might explain the decline of skill assessed at local level, especially over Europe, between lead time zero and lead time two (first column of Figure 6.4).

The third variability mode in July, which explains 10% of the total variance, also displays many features as its observed counterpart in May (top right panel of Figure 6.8). The expansion coefficient associated with this mode is only weakly correlated with the AO (-0.05) and Scandinavian (0.10) indices, the ones that have the highest correlation with the variability mode in May. On the other hand, it is negatively correlated with the EA index (-0.50). This correlation might be explained because the anomalies over Russia/Kazakhstan and over the North Atlantic off the cost of North Africa and the Iberian Peninsula resembles those of precipitation anomalies associated with the EA teleconnection pattern in July. The predicted patterns by individual forecast systems at lead time two differ from the observed one (left column of Figure 6.8). For instance, S4 predicts anomalies of opposite sign over the Norwegian Sea and Northern Africa.

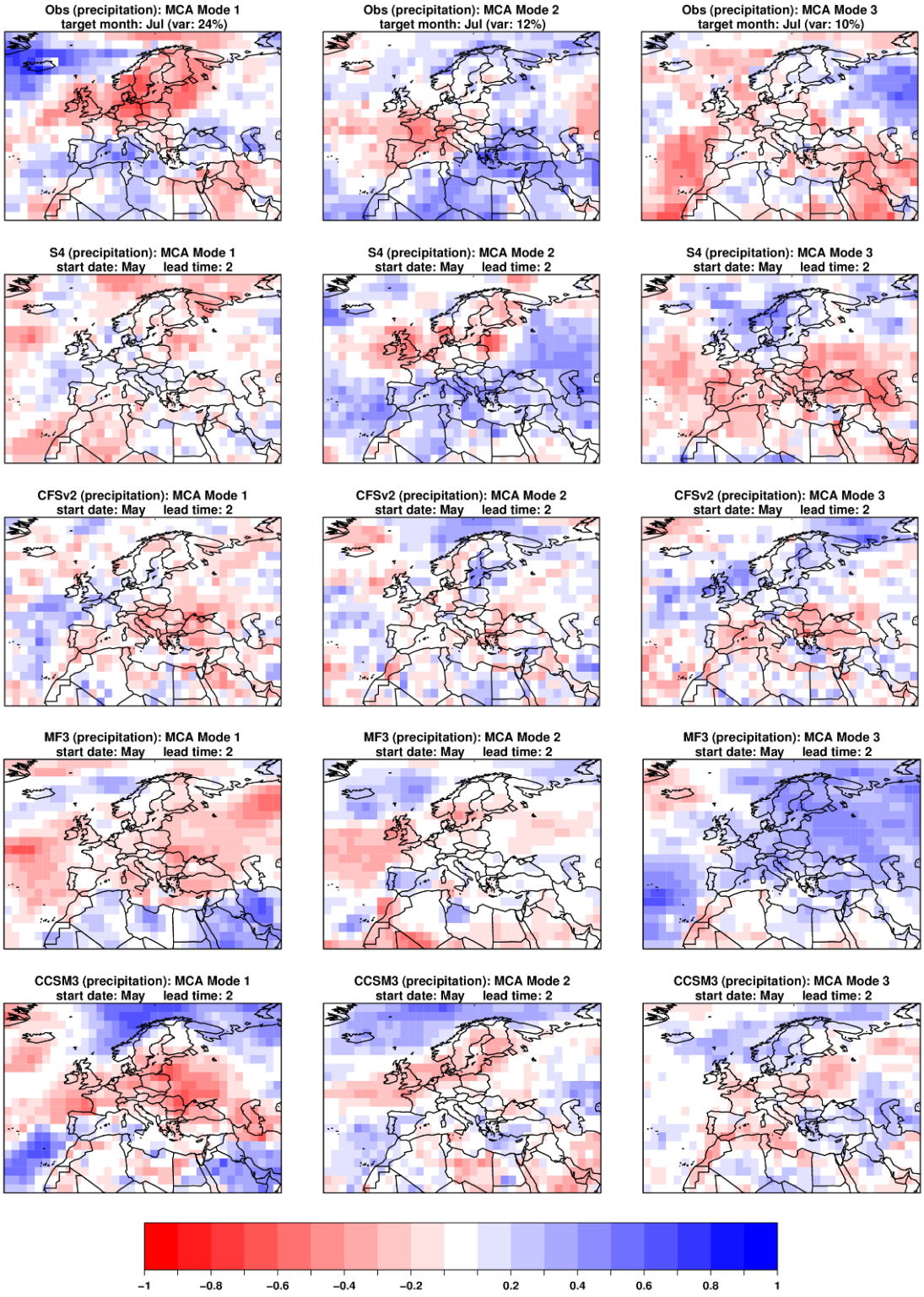

Figure 6.8: Observed and predicted heterogeneous correlation maps of precipitation in July. Predictions are for May start date (lead time 2). The expansion coefficients of the left field (i.e. the observation) are correlated with the original data of the right field (i.e. the forecast systems) and vice versa. Results are shown for the first three leading MCA modes, computed for the hindcast period 1982-2010.

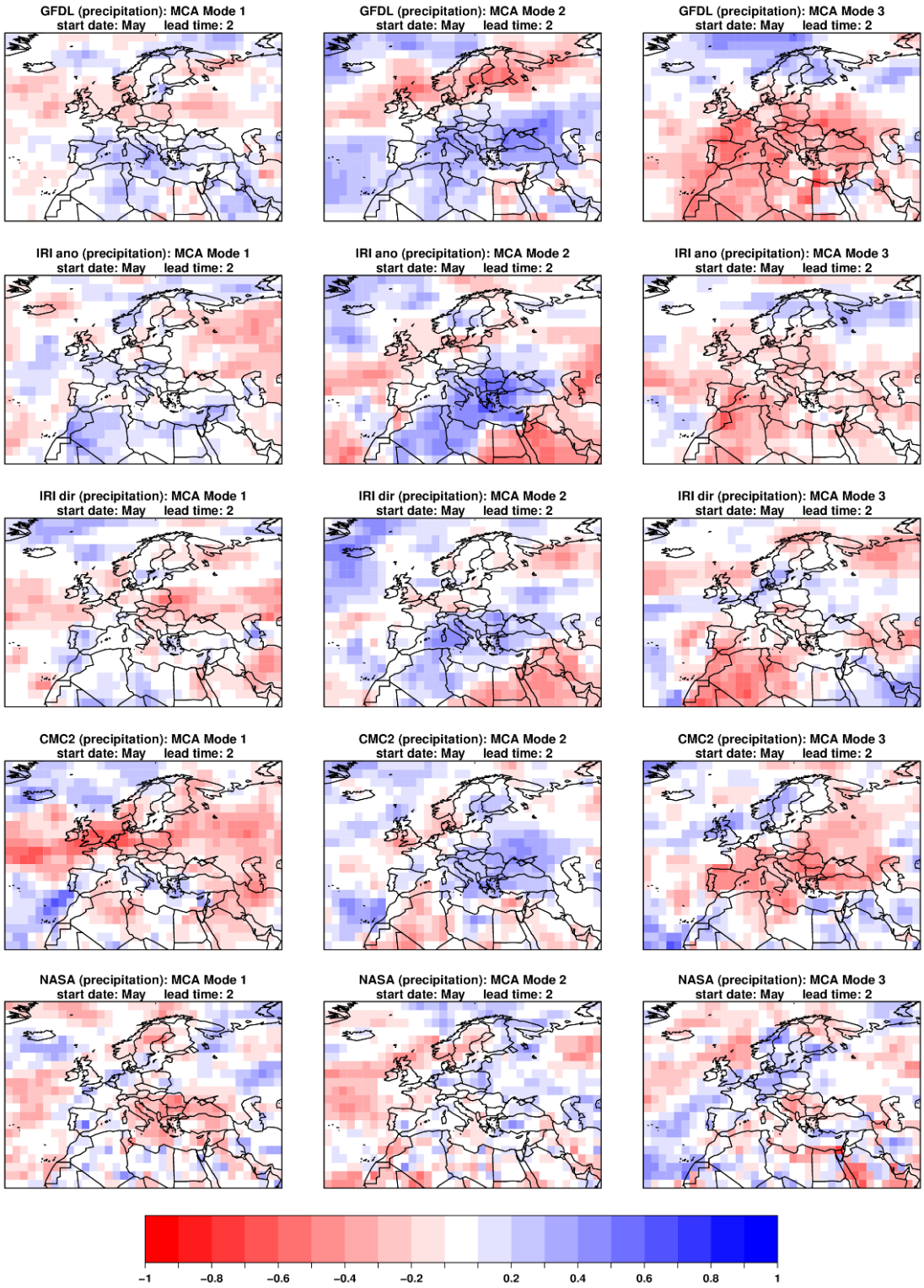

Figure 6.8: Continue

The first observed precipitation variability mode in December (predictions at lead time one month), which explains 31% of the total variance, displays widespread positive anomalies over Western Europe and negative anomalies south and north of it (top left panel of Figure 6.9). As previously, the expansion coefficient associated with this variability mode is correlated with several climate indices to attempt to find a physical

meaning it. Among several indices, it presents the highest correlation with the EA index (0.41). In turn, the precipitation teleconnection pattern in December associated with this index is very similar to the first observed precipitation variability mode in the same month. Most forecast systems detect a dipole of positive anomalies over Europe and negative anomalies south of it; however, all of them present limitations concerning the position and magnitude of the anomalies (left column of Figure 6.9). For instance, S4, CFSv2 and MF3 predictions display a dry Mediterranean whereas IRI-ECHAM direct present a wet eastern bound of the Mediterranean. This figure shows how difficult it is for current dynamical forecast systems to reproduce the observed precipitation variability in the extra-tropical regions in winter with lead time one month.

The second observed precipitation variability mode in December displays strong negative anomalies over northern North Atlantic, positive anomalies over Europe and Eurasia and again negative anomalies over North Africa and Middle East (top central panel of Figure 6.9). The expansion coefficient associated with this mode (explained variance 17%) is negatively correlated with the NAO (-0.60), AO (-0.43) and EAWR (-0.46). In fact, the second observed precipitation variability mode in December have many features found in the precipitation pattern associated with these three indices. For example, positive anomalies over northern Europe could be linked to the precipitation pattern of the positive phase of the NAO and AO associated with the strengthening of the zonal circulation (Figure 2.2). On the other hand, neither the negative anomalies over the Iberian Peninsula nor the positive anomalies over northeastern Africa associated with the NAO and AO pattern is observed in the top central panel of Figure 6.9. The precipitation pattern associated with the EAWR presents many similar features to the ones found in the second observed precipitation variability mode in December, including the strong negative anomalies over northern North Atlantic. IRI-ECHAM anomaly is the only forecast system that captures these strong anomalies over northern North Atlantic associated with the EAWR (central column of Figure 6.9). Similarly, CFSv2 and NASA are the only forecast systems that capture the positive anomalies associated with the NAO and AO precipitation patterns in December.

The third variability mode, which is explained by 9% of the total variance, displays negative anomalies over northern Europe and adjacent oceans and positive anomalies over the Iberian Peninsula and adjacent ocean (top right panel of Figure 6.9). The ocean component of this dipole is similar to the precipitation pattern associated with the EAWR in December, which could explain the correlation between the expansion coefficient associated with this variability mode and the EAWR index (-0.34). Most forecast systems fail to simulate this variability mode, with the exception of CMC2 that captures the ocean component of the above-mentioned precipitation dipole over North Atlantic.

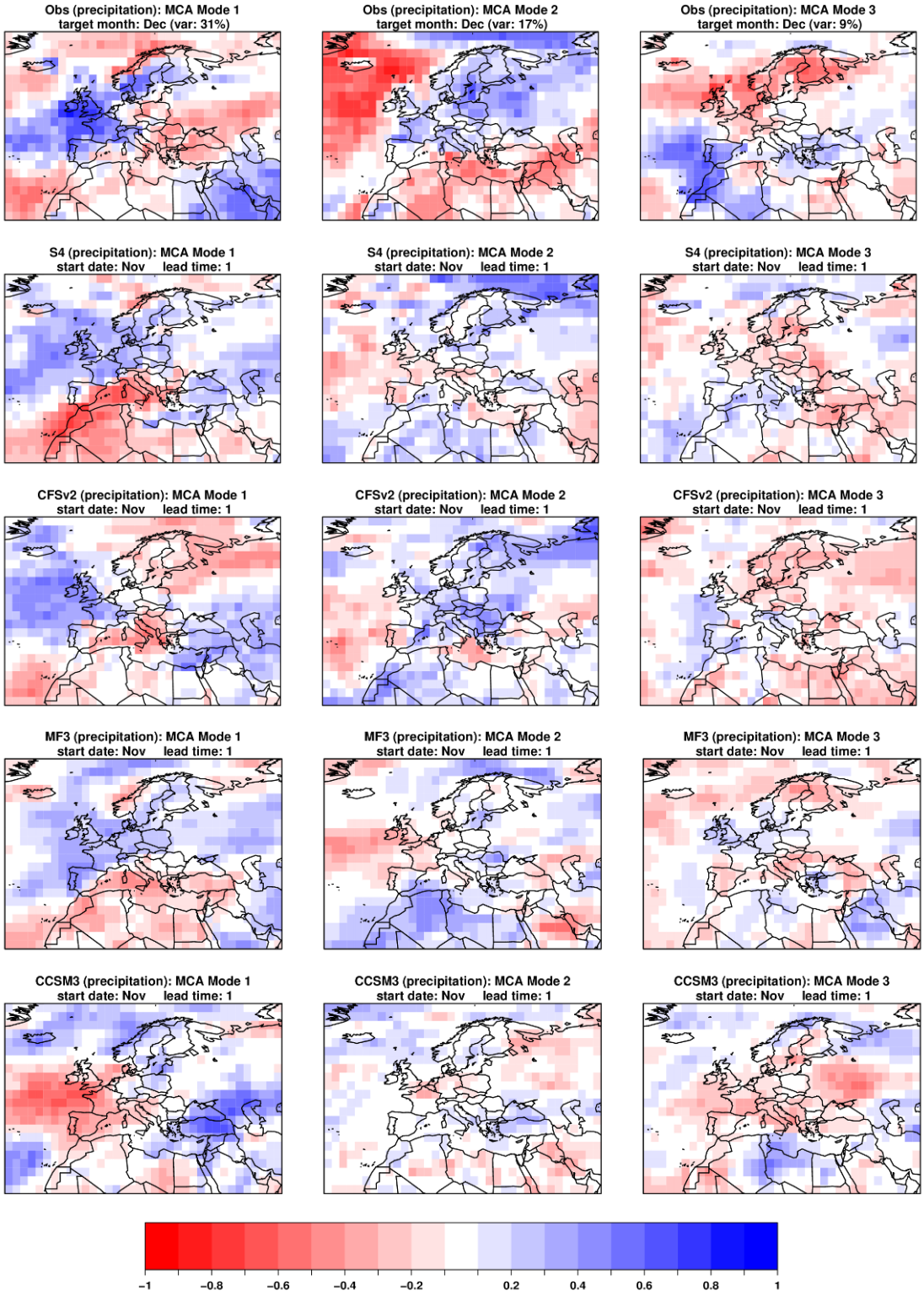

Figure 6.9: Observed and predicted heterogeneous correlation maps of precipitation in December. Predictions are for November start date (lead time 1). The expansion coefficients of the left field (i.e. the observation) are correlated with the original data of the right field (i.e. the forecast systems) and vice versa. Results are shown for the first three leading MCA modes, computed for the hindcast period 1982-2010.

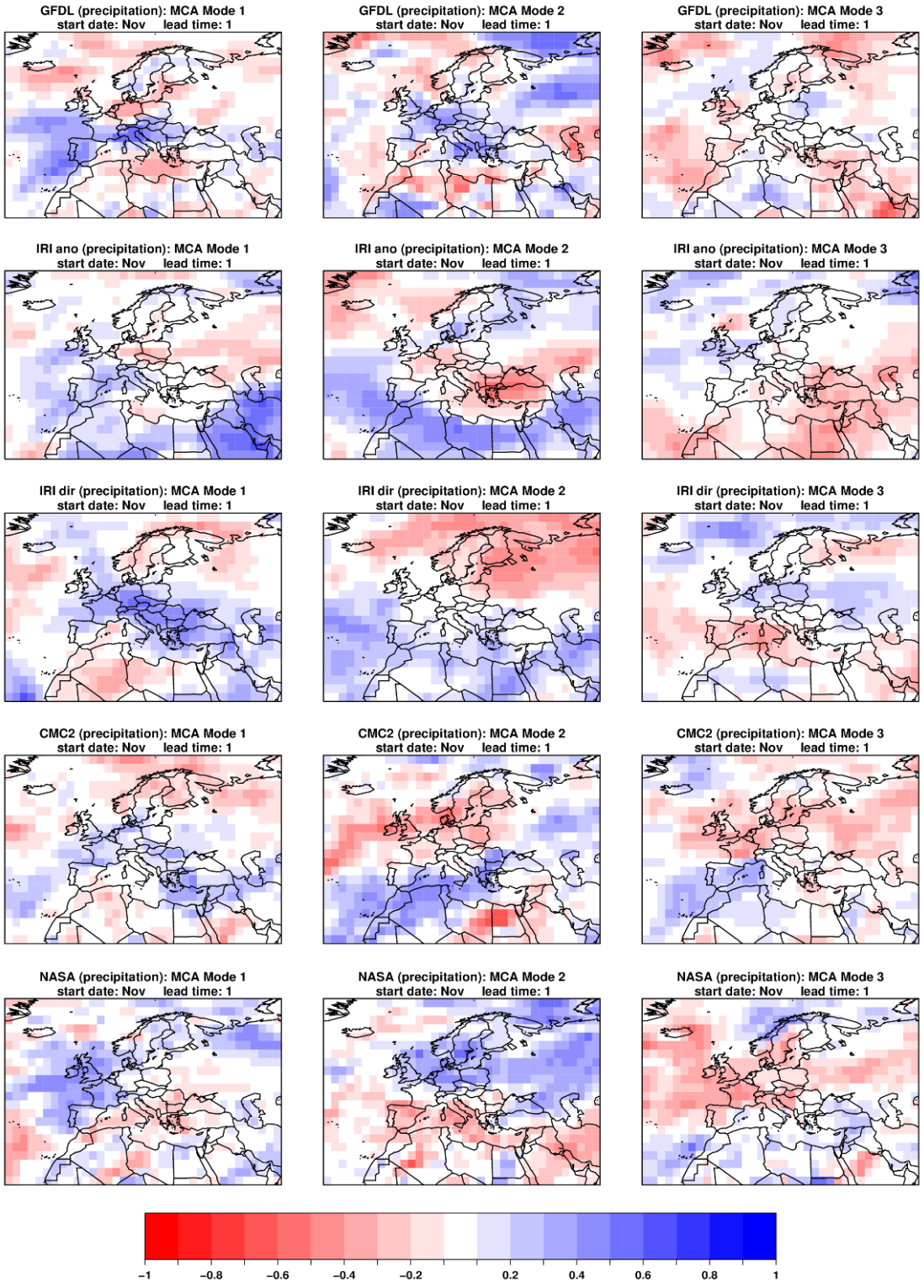

Figure 6.9: Continue.

The expansion coefficients associated with the MCA modes of variability are used to estimate the parameters needed to combine several models using the FA method. Figure 6.10 illustrates the expansion coefficients associated with the first three modes of near-surface temperature variability in June (top left panel of Figure 6.6). The expansion coefficient associated with the first near-surface temperature variability mode displays a

positive trend over the studied region, which is consistent with the corresponding spatial pattern as most of the grid points have the same sign. This is consistent with the observed warming trend over Europe shown in Figure D.1, a feature already documented previously (EEA, 2015). On the other hand, both the second and the third PCs display interannual variability. The expansion coefficients associated with the first three precipitation variability modes in December (predictions at lead time one month) also displays an interannual variability (second row of Figure 6.10).

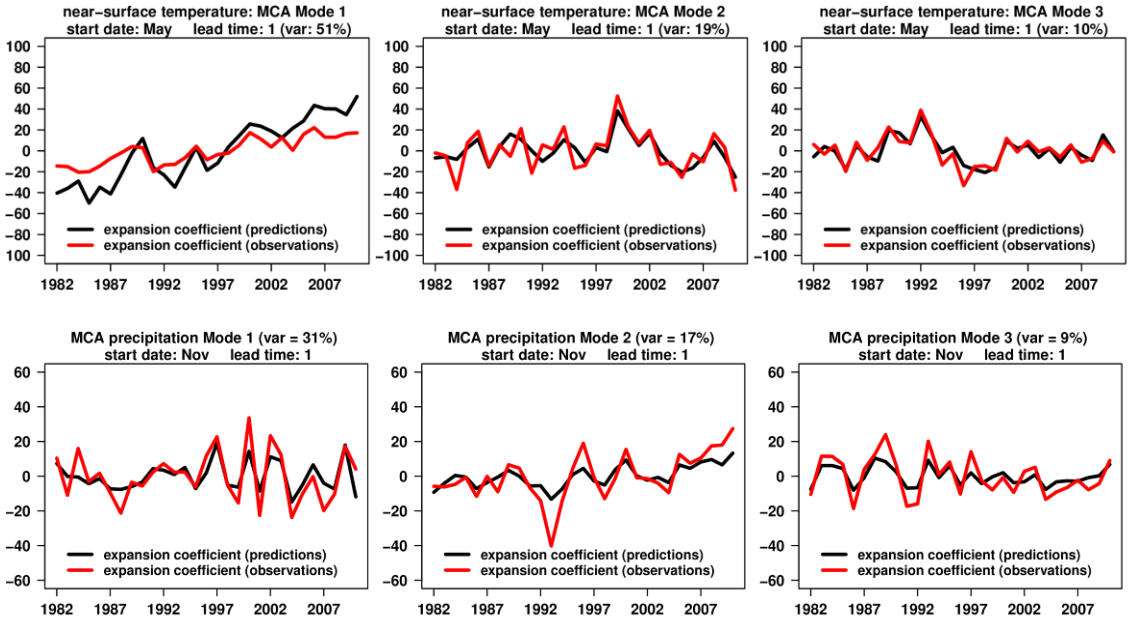

Figure 6.10: First three expansion coefficients of left (observations) and right (predictions) fields for near-surface temperature in June (top row) and precipitation in December (bottom row). Forecasts are for May starting date (lead time 1).

After estimating the MCA modes to reduce the dimension of the original dataset, an important question that arises is how many expansion coefficients would be retained to lead to the best FA prediction. As in Coelho et al. (2006), we have retained and tested a range of number of modes (i.e. the FA predictions are estimated with two, three, four, five and six modes). The number of modes that give the best FA prediction varies according to the climate variable, start date, lead time and area. However, the FA predictions estimated using three MCA modes give the best prediction more often than not, both for precipitation and near-surface temperature.

6.3.2. Choice of the number of MCA modes used in the FA combination

Figure 6.11 shows the correlation between the observations and the FA predictions, estimated using a different number of MCA modes in three-year-out cross-validation mode (i.e. the target year, one year prior to and one year after the target year are removed from the analysis). Predictions were initialized in May to predict May, June, July and August (from the left to the right column). FA predictions estimated using two MCA modes are displayed in the first row, the ones using three MCA modes in the second row, and so forth. FA predictions at lead time 0 improve as more MCA modes are included in

the analysis; however, only little improvement is found when the sixth mode is added (left column of Figure 6.11). On the other hand, FA predictions using three modes performs best when predicting with lead times one, two and three months, especially over Europe and northern Africa. This might be explained because the first MCA variability mode (i.e. the warming trend) explains only 33% of the cross-covariance matrix of the predictions at lead time 0 and the observations. This variance increases to 51%, 56% and 66% for predictions at lead times 1, 2 and 3, respectively, because the predicted warming trend increases with lead time probably due to systematic errors. As a result, FA predictions that takes into account the variability mode that are best predicted by most forecast systems (first column of Figure 6.6) generate the most skillful forecasts. On the other hand, FA performs badly over the Eurasia region where forecast systems fail to predict the weak negative anomalies in the first variability mode as well as the strong negative anomalies in the second variability mode (first column of Figure 6.6).

6.3.3. Forecast quality of the combinations

The resulting skill from combining multiple forecast systems is illustrated and discussed in this subsection. The SMM and the FA methods are used to combine all dynamical forecast systems shown above as well as only the four best systems (i.e. the ones that perform better more often). The best forecast systems are S4, CFSv2, GFDL and CMC2. The aim is to quantify the relative merits of adding less skillful forecast systems in the resulting combined prediction. For simplicity, statistical models are not included in the combination. Instead, a climatological PDF is used as the prior prediction in the FAC combination.

The correlation coefficient between predicted and observed near-surface temperature in June (predictions at lead time one month) illustrates the deterministic skill (correlation of the ensemble mean) of the four combinations used in this study Figure 6.12). The correlation of the SMM and FAC predictions displays an overall similar pattern, but with significant differences at local level in some regions. For instance, FAC outperforms SMM over the Scandinavian Peninsula, but the opposite happens over the Eurasian region. The weak performance of the FAC over the Eurasian region might be explained because most forecast systems fail to predict weak (strong) negative anomalies in the first (second) near-surface temperature variability mode (first and second columns of Figure 6.6). The poor performance of FAC is observed in target months and lead times over regions where the forecast systems fail to predict the main variability modes (Appendices F and G).

The combination of a different number of forecast systems using the same methodology shows that the addition of more forecast systems does not always lead to better forecast skill, especially when they lack skill. For example, the FA over Russia performs worse when all forecast systems are used in the combination (third and fourth panels of Figure 6.12). Independently from the combination method, near-surface temperature correlation skill is more sensitive to lead time in winter (Figure F.2 and G.2) than in summer (Figure F.1 and G.1), probably because much of the skill in summer derived from the trend as previously described in the Section 6.3.1. Another important result is that the best forecast system (i.e. S4) generally performs at least as good or better than the best resulting

combination (i.e. SMM with S4, CFSv2, GFDL and CMC2). This can be seen by comparing the third column of Figure 6.4 with the Figure F1 and the fourth column of Figure 6.4 with the Figure F2, both Figures of the Appendix F. However, it is important to remember that the best forecast system is not always the same (Hagedorn et al., 2005). For instance, CFSv2 is the best forecast system over many regions when predicting near-surface temperature in June with lead time one month (Figure 6.2). For this specific prediction, the SMM (second column of Figure 6.12) outperforms S4 (top left column of Figure 6.2) in many grid-points over in the Eurasia region. In Chapter 5, it was shown that S4 alone predicts the modes of precipitation variability over the WAM region better than all other single forecast system and their combination more often than not.

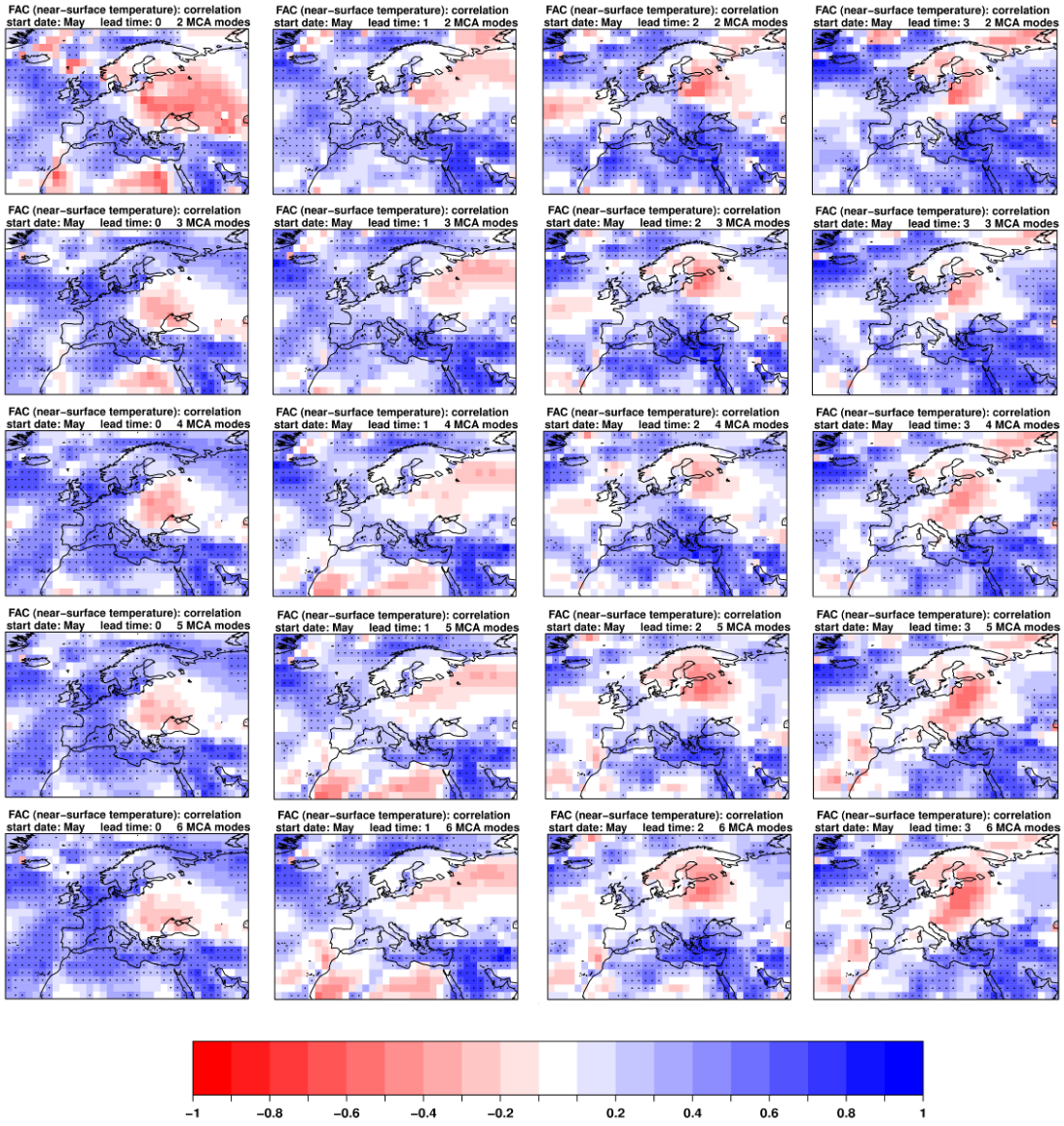

Figure 6.11: Correlation between the observed and FA predictions of near-surface temperature in May, June, July and August (from left to right). Predictions are for May start date. FA predictions are estimated using two, three, four, five and six MCA modes (from top to bottom). The correlation was computed for the hindcast period 1982-2010. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

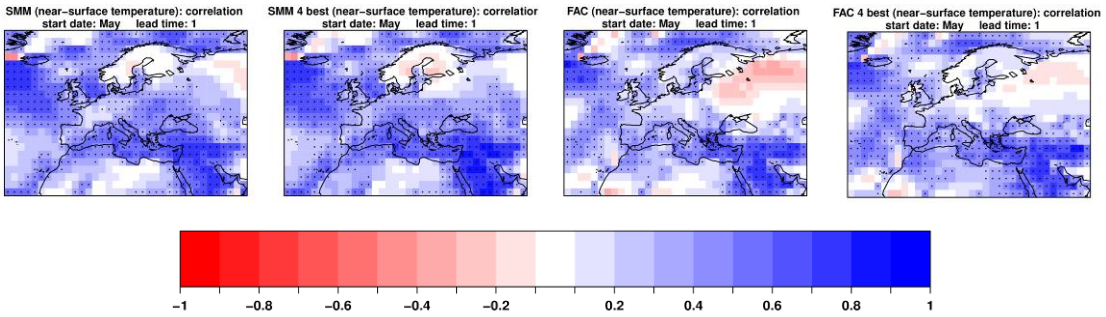

Figure 6.12: Correlation coefficient between predicted and observed near-surface temperature in June. Predictions are for May start dates (lead time 1). The correlation was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

SMM outperforms FAC when predicting precipitation more often than not (Appendix F). However, there are many illustrations when the FAC precipitation predictions perform better. For instance, FAC 4 best (fourth column of Figure 6.13) performs better than SMM (first and second column of Figure 6.13) over the northern North Atlantic, west the cost of Great Britain, and over parts of Russia. In this illustration, the skill pattern is similar in both combination methods. Two conclusions hold the analysis of the combination for near-surface temperature predictions. First, forecast skill does not improve with the addition of more forecast systems to the combination, especially when it is known that they lack skill (Figures 6.13). Here the forecast systems not considered in the 4 best combinations are MF3, CCSM3, IRI-ECHAM systems and NASA. This directs attention to the importance of continuously assess the forecast quality for specific situations. Second, S4 outperforms the other forecast systems and their combination more often than not in terms of correlation coefficient (not shown). For the precipitation predictions in December with lead time one month (Figures 4 and 10), the SMM 4 best outperforms performs any single forecast system more often than not.

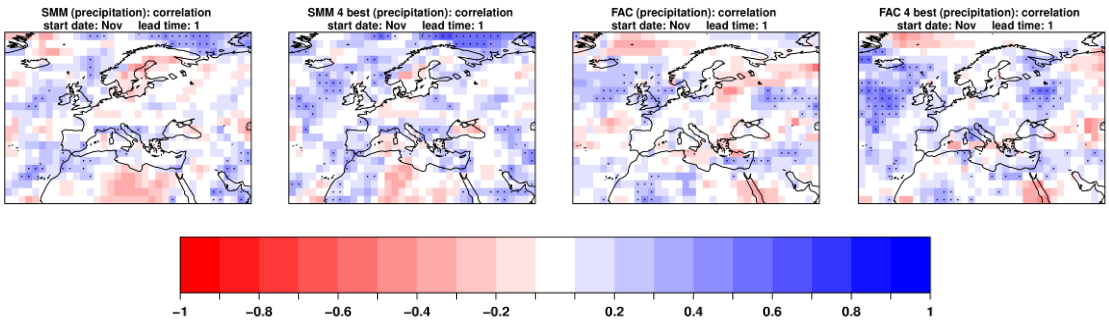

Figure 6.13: Correlation coefficient between predicted and observed precipitation in December. Predictions are for November start dates (lead time 1). The correlation was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

Several probabilistic scores are estimated to better quantify the benefits and limitations of combining several forecast systems using the two different combination methods considered in this chapter. The CRPSS illustrates the probabilistic forecast quality of the combinations for the near-surface temperature predictions in June with May start date (Figure 6.14). For this specific illustration, the patterns of the forecast skill of the SMM and FAC are similar with statistically significant positive CRPSS over northern North Atlantic, Mediterranean Sea (limited to the eastern bound in the FA predictions) and Middle East. However, the FAC probabilistic predictions are generally reduced when compared to SMM in the other target months and lead times (not shown). One of the reasons that might explain the weaker performance of the FA method compared to SMM is associated with the need of the former for dimension reduction. Following Stephenson et al. (2005), MCA is used in this chapter to reduce the dimension of the data to a few variability modes that explains most of the variance. This approach proved good in regions where the main variability modes are successfully predicted by current forecast systems, such as regions dominated by the ENSO signal (Coelho et al., 2006). However, the main near-surface temperature and precipitation variability modes over most extratropical regions are poorly predicted by current forecast systems as shown in the Section 6.3.1.

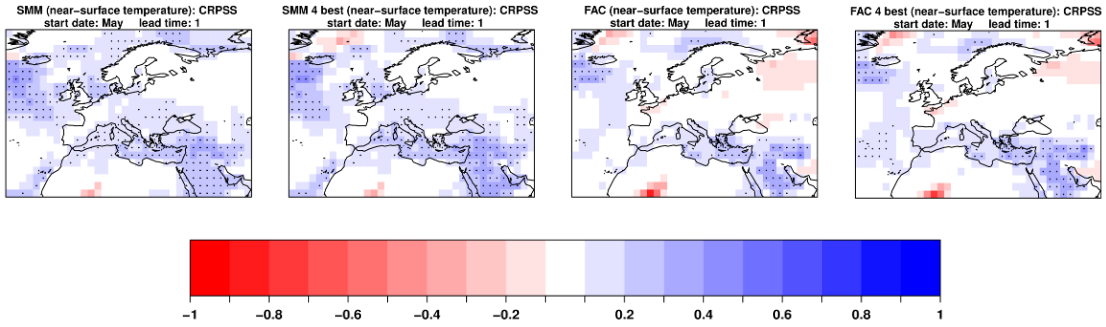

Figure 6.14: CRPSS for near-surface temperature predictions in June. Predictions are for May start dates (lead time 1). The CRPSS was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

6.4. Summary and conclusions

This study performed a forecast quality assessment of monthly near-surface temperature and precipitation predictions performed by simple statistical models and dynamical forecast systems over Europe. Operational and quasi-operational dynamical forecast systems from both the European EUROSIP initiative and North American NMME project are used. The dynamical forecast systems are combined using two different combination methods: one with equal weights, assuming all forecast systems have similar performance, and one giving more weights to the forecast systems that perform better over the hindcast period. The predictions of the resulting combinations are assessed and compared to the predictions of single forecast systems. Two start dates (May and

November) and four lead times (zero through three) are considered to take into account four target months in two seasons: summer (MJJA) and winter (NDJF).

Simple linear regression models are estimated considering several predictors: the predictand itself and three SST indices (Niño3.4, SNA and AMO). These models are used as reference forecasts. Therefore, predictors of the month of April are used to predict near-surface temperature and precipitation in May, June, July and June and predictors of the month of October are used to predict these variables in November December, January and February, allowing a fair comparison with the dynamical forecast systems. The parameters of the statistical models are estimated in retroactive mode, that is, only the period prior to the target period is used to estimate the regression coefficient emulating an operational prediction. Consequently, datasets with long-term record, available only over the continents, are used to estimate these parameters. For simplicity, a combination of predictors (e.g. multiple linear regression) has not been taken into account. We have shown that the forecast skill of the statistical models does not change significantly with lead time, in agreement with previous studies (e.g. Barnston, 1994; Lang et al., 2014). Regardless of the predictor, simple linear regression models are more skillful when predicting near-surface temperature than precipitation. Besides, near-surface temperature predictions are more skillful in summer than in winter, which might be linked the observed warming trend in summer. A recent study has shown that most of the forecast skill of near-surface temperature predictions vanishes and gets limited to the ENSO teleconnection regions when the warming trend is removed prior to the analysis (Eden et al., 2015). None of the four statistical models used here provide skillful precipitation predictions over Europe, with the exceptions for very few grid points with significant positive correlation.

The forecast quality assessment of the dynamical forecast systems is also analyzed. As with the statistical models, dynamical forecast systems predict better near-surface temperature than precipitation, which might be linked to the fact that most forecast systems predict the warming trend over the studied regions although at a different rate when compared to the observation. On the other hand, differently from the statistical model predictions, forecast skill varies considerably with lead time and forecast system. For example, while S4 and CFSv2 are found to perform better than the IRI-ECHAM systems, statistical models built with different predictors present similar skill patterns. This can be seen by comparing Figure 6.1 with the Figures 6.2 and 6.3. At the same time, the S4 precipitation predictions loose considerable amount of skill from lead time 0 to lead time 1, both in summer and winter (Figure 6.4).

The FA method is used to combine the dynamical forecast systems assigning more weight to systems that presented better performance over the hindcast period. The FA method has been implemented using a prior taken from the climatology, the FAC. The skill of the resulting FAC predictions is compared to the skill of the SMM combination where forecast systems are combined without assigning weights. The FAC generally presents reduced scores when compared to SMM for near-surface temperature and precipitation over Europe, both in terms of deterministic and probabilistic verification measures. However, there are many instances when the spatial pattern of the FAC forecast skill is similar to that of the SMM approach (Figure 6.12, 6.13, 6.14). One of the reasons that

might explain the reduced forecast skill of the FAC predictions for specific regions over Europe is associated with the way the coefficients are estimated. The first step to estimate the FAC predictions is to apply a dimension reduction technique like the MCA to deal with the high dimensionality of the gridded data and the strong dependency between values of neighboring grid points (Stephenson et al., 2005; Coelho et al., 2006). However, as shown in the Section 6.3.1, the forecast systems are not able to properly simulate the main modes of near-surface temperature and precipitation variability over the extratropics. An illustration of how the predicted variability modes affect the skill of the FAC predictions can be seen by comparing Figure 6.6 with Figures 6.12 and 6.14. In this illustration, there is a large area where FAC predictions are negatively correlated with the observations over Russia, a region where most forecast systems do not predict the correct sign of the anomalies in the first and second observed variability mode. Previous studies have shown that the FAC predictions perform at least as well as the SMM in regions where the variability modes are well predicted by forecast systems, particularly the ones affected by ENSO (e.g. Coelho et al., 2006).

Seasonal forecast skill relies on the response of the atmosphere to the slowly varying components of the climate systems. This response is not well quantified by currently forecast systems over certain regions across the globe, including over Europe. As a result, forecast systems have little skill over Europe. In this chapter, we showed that precipitation forecast skill considerably decreases from lead time zero to lead time one month. On the other hand, forecast systems simulate well the summer near-surface temperature trend, although at a different rate when compared to the observed one, capturing the response of the atmosphere to the increasing greenhouse gases. The next generation of forecast systems need to quantify this response in order to improve forecast skill at seasonal timescales. This challenge has already been addressed by leading climate prediction centers (e.g. Scaife et al., 2014).

7. General Conclusions

The main objective of this thesis is twofold: first, to apply different statistical methods to combine seasonal predictions of climate variables produced by the state-of-the-art statistical and dynamical forecast systems; second, to assess the forecast quality of the resulting combination as well as predictions produced by the individual forecast systems. This thesis focuses on seasonal climate predictions of SST and precipitation in the tropics and near-surface temperature and precipitation in the extratropics.

The conclusions derived from this thesis are summarized below:

- Many of the dynamical forecast systems used in this study have skill when predicting climate variables over tropical regions. In Chapter 4, it is shown that three dynamical forecast systems (i.e. S4, CFSv2 and MF3) have statistically significant deterministic and probabilistic skill during all months of the year for predictions of SST indices over three tropical ocean basins: Pacific, Atlantic and Indian Oceans. In many instances, forecast skill is found up to lead time six months, the longest lead time available for S4 and MF3. In Chapter 5, many forecast systems have deterministic forecast skill when predicting the modes of WAM rainfall variability. However, only one of them (i.e. S4) produces skillful probabilistic forecasts, assessed in terms of CRPSS and ignorance skill score.

- On the other hand, forecast systems have low seasonal forecast skill over Europe and adjacent areas. In Chapter 6, it is illustrated that precipitation forecasts hardly have any statistically significant positive correlation over Europe in December for predictions produced in November (lead time one month). Near-surface temperature forecast have statistically significant correlation in summer months for predictions with lead time up to four months. However, most of the correlation skill might be explained because many forecast systems are able to reproduce the observed warming trend in summer months over most of Europe. None of the forecast systems produces statistically significant positive CRPSS in June with lead time one month.

- Simple linear regression models, used here as a benchmark, outperformed many of the dynamical forecast systems in several instances. This is illustrated for deterministic and probabilistic predictions of the WTI SST index in summer months and the modes of the WAM rainfall variability also in summer. However, there are occasions when simple linear models are of limited usefulness, such as to predict precipitation over Europe and adjacent areas.

- It is shown that no single forecast system performs best in all cases. For example, S4 is the best forecast system when predicting the modes of WAM rainfall variability and Niño3.4 SST index, CFSv2 performs best when predicting SNA SST index and the simple linear regression model performs best when predicting the WTI SST index.

- The predictions of the SMM are often better than those combination methods that assign unequal weights. The difficulty in the robust estimation of the weights due to the small samples available is one of the reasons that limit the potential benefit

of the combination methods that assign unequal weights. However, some of the results shown here give light to further research on how to improve the SMM predictions using combination methods that assign unequal weights. Two cases can illustrate this as follows:

1. Predictions of univariate predictands: the combination methods that assign unequal weights improve the SMM predictions when only a fraction of all single forecast systems have skill as in the case of the SNA index predictions in the boreal fall (Figure 4.5) or in the case WTI index predictions in the boreal summer (Figure 4.7). Therefore, the weighting does not outperform the SMM when the SMM is very skillful, but it reduces the risk of low skill situations that are found when several single forecast systems have a low skill.

2. Predictions of multivariate predictands: the FA method performs as well as or better than SMM in regions where the forecast systems used in the combination procedure are able to represent the modes of climate variability associated to predictive skill. However, this does not hold true over most of Europe and adjacent regions. In these cases, not only SMM performs better than FA, but also SMM is frequently outperformed by the best single forecast system (S4). After building a multimodel ensemble with a smaller number of forecast systems, it is shown that the addition of more forecast system does not always lead to better forecast.

- There are cases when combining all forecast systems does not lead to improved forecasts when compared to the best single forecast system. In fact, S4 is far better than any forecast system when predicting the variability of the WAM rainfall regimes several months ahead (Figure 5.9 and 5.10). This suggests that in some special occasions like this one, a multimodel approach is not necessarily better than an especially skillful forecast system. This shows the importance of continuously assessing the forecast quality for the specific application of the user.

This study has applied and compared several combination methods, which had been previously described in the literature. These methods were used to combine climate predictions produced by several operational and quasi-operational forecast systems, most of which are publicly available on the web or easily implemented by the user. This allowed a uniform assessment (i.e. using the same verification measures and observations) of the predictions produced by each combination method and each single forecast system. However, due to the limited time available for a PhD research, many research questions were not dealt with in this PhD thesis, including some that were formulated during this PhD thesis. Some of these questions could be recommended for future studies in this field of research as follows:

- The forecast quality of several dynamical forecast systems, each having a different number of ensemble members, were assessed without asking the question of whether or not or to what extent the number of ensemble members affect the forecast quality of the forecast system. That is, if a certain forecast system is better than another one simply because it has a larger number of ensemble member or if

it actually has a better representation of the climate system. This kind of issue could be dealt with by using the so-called *fair scores* (Ferro, 2014).

- Another issue found in this study was to compare a probabilistic prediction estimated from an ensemble forecast system using a frequentist approach (i.e. number of ensemble members above or below a certain threshold) with a probabilistic prediction derived from a predictive distribution function. One way to address this issue could be to estimate a predictive distribution function from an ensemble system before estimating the probabilistic prediction.

- It is shown in Chapter 6 that near-surface temperature predictions over Europe and adjacent regions have skill in summer months in terms of correlation coefficient, but not in terms of CRPSS. This might be linked to the fact that many forecast systems simulate the observed warming trend (Appendix D). However, a quantification of how much of the correlation skill comes from the warming trend has not been investigated here. This could be studied by assessing the forecast quality of near-surface temperature after the trend is removed.

- It is shown in Chapter 6 that one of the challenges of using the FA over Europe and adjacent regions is associated with the fact that current forecast systems do not simulate well the modes of near-surface temperature (except the first mode, which is the warming trend) and precipitation variability. However, this has not been quantified. To address this issue, one could perform a forecast quality assessment of these variability modes for each single forecast system as it was done in Chapter 4 for the modes of WAM rainfall variability.

- This thesis has focused on SST, near-surface temperature and precipitation. However, the uncertainty quantification and forecast quality assessment performed in this study might be also of value to other surface climate variables such as wind, especially in the context of renewable energy.

References

Alessandri, A., A. Borrelli, A. Navarra, A. Arribas, M. Déqué, P. Rogel, A. Weisheimer (2011) Evaluation of Probabilistic Quality and Value of the ENSEMBLES Multimodel Seasonal Forecasts: Comparison with DEMETER, *Monthly Weather Review*, 139, 581-607

Arguez, A., R. S. Vose (2011) The definition of the standard WMO climate normal: The key to deriving alternative climate normals, *Bulletin of the American Meteorological Society*, 92, 699-704

Arribas, A., M. Glover, A. Maidens, K. Peterson, M. Gordon, C. MacLachlan, R. Graham, D. Fereday, J. Camp, A. A. Scaife, P. Xavier, P. McLean, A. Colman, S. Cusack (2011) The GloSea4 ensemble prediction system for seasonal forecasting, *Monthly Weather Review*, 139, 1891-1910

Balmaseda, M. A., M. K. Davey, D. L. T. Anderson (1995) Decadal and seasonal dependence of ENSO prediction skill, *Journal of Climate*, 8, 2705-2715

Barnett, T. P., R. Preisendorfer (1987) Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis, *Monthly Weather Review*, 115, 1825-1850

Barnston, A. G., H. M. Van den Dool (1993) A degeneracy in cross-validated skill in regression-based forecasts, *Journal of Climate*, 6, 963-977

Barnston, A. G. (1994) Linear statistical short-term climate predictive skill in the Northern Hemisphere, *Journal of Climate*, 7, 1513-1564

Batté, L., M. Déqué (2011) Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: skill of the ENSEMBLES project multimodel ensemble forecasts, *Tellus A*, 63, 283-299

Bengtsson, L., U. Schlese, E. Roeckner, M. Latif, T. P. Barnett, N. Graham (1993) A two-tiered approach to long-range climate forecasting, *Science*, 261, 1026-1029

Bjerknes, V. (1904) Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik (The problem of weather prediction, considered from the viewpoints of mechanics and physics), *Meteorologische Zeitschrift*, 21, 1-7. Translated and edited by E. Volken and S. Bronnimann (2009), *Meteorologische Zeitschrift*, 18, 663-667

Bjerknes, J. (1966) A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature, *Tellus*, 18, 820-829

Bjerknes, J. (1969) Atmospheric teleconnections from the equatorial Pacific, *Monthly Weather Review*, 97, 163-172

Blender, R., U. Luksch, K. Fraedrich, C. C. Raible (2003) Predictability study of the observed and simulated European climate using linear regression, *Quarterly Journal of the Royal Meteorological Society*, 129, 2299-2313

Bouali, L., N. Philippon, B. Fontaine, J. Lemond (2008) Performance of DEMETER calibration for rainfall forecasting purposes: Application to the July-August Sahelian rainfall, *Journal of Geophysical Research*, 113, D15111

Brönnimann, S. (2007) Impact of El Niño–Southern Oscillation on European climate, *Reviews of Geophysics*, 45, 1-36

Buizza, R., M. Milleer, T. N. Palmer (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, *Quarterly Journal of the Royal Meteorological Society*, 125, 2887-2908

Charney, J. G., J. Shukla (1981) Predictability of monsoons, *Monsoon dynamics*, 99-109

Chase, T. N., R. A. Pielke, R. Avissar (2007) Teleconnections in the earth system, *Encyclopedia of Hydrological Sciences*, 15, 183

Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, D. B. Stephenson (2004) Forecast calibration and combination: A simple Bayesian approach for ENSO, *Journal of Climate*, 17, 1504-1516

Coelho, C. A. S., D. B. Stephenson, M. Balmaseda, F. J. Doblas-Reyes, G. J. van Oldenborgh (2006) Toward an integrated seasonal forecasting system for South America, *Journal of Climate*, 19, 3704-3721

Coelho, C. A. S., S. M. Costa (2010) Challenges for integrating seasonal climate forecasts in user applications, *Current Opinion in Environmental Sustainability*, 2, 317-325

Cohen, J. (2011) Eurasian snow cover variability and links with stratosphere-troposphere coupling and their potential use in seasonal to decadal climate predictions

Cohen, J., J. Jones (2011) A new index for more accurate winter predictions, *Geophysical Research Letters*, 38, L21701

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A. J. Weaver, M. Wehner (2013) Long-term Climate Change: Projections, Commitments and Irreversibility. In*: Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Cook, K. H., E. K. Vizy (2006) Coupled model simulations of the West African monsoon system: Twentieth- and twenty-first-century simulations, *Journal of Climate*, 19, 3681-3703

Curry, J. A., P. J. Webster (2011) Climate science and the uncertainty monster, *Bulletin of the American Meteorological Society*, 92, 1667-1682

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Kohler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thepaut, F. Vitart (2011) The ERA-Interim reanalysis: configuration

and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553-597

DelSole, T., X. Yang, M. K. Tippett (2012) Is unequal weighting significantly better than equal weighting for multi-model forecasting?, *Quarterly Journal of the Royal Meteorological Society*, 139, 176-183

DeWitt, D. G. (2005) Retrospective Forecasts of Interannual Sea Surface Temperature Anomalies from 1982 to Present Using a Directly Coupled Atmosphere-Ocean General Circulation Model, *Monthly Weather Review*, 133, 2972-2995

Dima, I. M., J. M. Wallace (2003) On the Seasonality of the Hadley Cell, *Journal of the Atmospheric Sciences*, 60, 1522-1527

Doblas-Reyes, F. J., M. Déqué, J. P. Piedelievre (2000) Multi-model spread and probabilistic seasonal forecasts in PROVOST, *Quarterly Journal of the Royal Meteorological Society*, 126, 2069-2087

Doblas-Reyes, F. J., V. Pavan, D. B. Stephenson (2003) The skill of multi-model seasonal forecasts of the wintertime North Atlantic Oscillation, *Climate Dynamics*, 21, 501-514

Doblas-Reyes, F. J., R. Hagedorn, T. N. Palmer (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination, *Tellus A*, 57, 234-252

Doblas-Reyes, F. J., R. Hagedorn, T. N. Palmer, J. J. Morcrette (2006) Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts, *Geophysical Research Letters*, 33, L07708

Doblas-Reyes, F. J., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. M. Murphy, P. Rogel, D. Smith, T. N. Palmer (2009) Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts, *Quarterly Journal of the Royal Meteorological Society*, 135, 1538-1559

Doblas-Reyes, F. J., J. Garcia-Serrano, F. Lienert, A. Pinto-Biescas, L. R. L. Rodrigues (2013a) Seasonal climate predictability and forecasting: status and prospects, *WIREs Climate Change*, 4, 245-268

Doblas-Reyes, F. J., I. Andreu-Burillo, Y. Chikamoto, J. García-Serrano, V. Guemas, M. Kimoto, T. Mochizuki, L. R. L. Rodrigues, G. J. Van Oldenborgh (2013b) Initialized near-term regional climate change prediction, *Nature Communications*, 4, 1715

Eden, J. M., G. J. van Oldenborgh, E. Hawkins, E. B. Suckling (2015) A global empirical system for probabilistic seasonal climate prediction, *Geoscientific Model Development Discussions*, 8, 3941-3970

Edwards, P. N. (2000) A brief history of atmospheric general circulation modeling, *International Geophysics Series*, 70, 67-90

Edwards, P. N. (2011) History of climate modeling, *Wiley Interdisciplinary Reviews: Climate Change*, 2, 128-139

EEA (2015) European Environment Agency: Global and European temperatures (CSI 012/CLIM 001) Assessment, downloaded from http://www.eea.europa.eu/data-and-maps/indicators/global-and-european-temperature-1/assessment, accessed at 7 August 2015

Enfield, D. B., A. M. Mestas-Nuñez, D. A. Mayer, L. Cid-Cerrano (1999) How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures?, *Journal of Geophysical Research*, 104, 7841-7848

Fan, Y., H. van den Dool (2008) A global monthly land surface air temperature analysis for 1948-present, *Journal of Geophysical Research*, 113, D01103

Feddersen, H., U. Andersen (2005) A method for statistical downscaling of seasonal ensemble predictions, *Tellus A*, 57, 398-408

Ferrel, W. (1859) The motions of fluids and solids relative to the earth's surface, *Mathematics Monthly*, 1, 140-148, 210-216, 300-307, 366-373, 397-406

Ferro, C. A. T. (2014) Fair scores for ensemble forecasts, *Quarterly Journal of the Royal Meteorological Society*, 140, 1917-1923

Folland, C. K., T. N. Palmer, D. E. Parher (1986) Sahel rainfall and worldwide sea surface temperature, *Nature*, 320, 602-607

Fontaine, B., S. Janicot, V. Moron (1995) Rainfall anomaly patterns and wind field signals over West Africa in August (1958-1989), *Journal of Climate*, 8, 1503-1510

Fontaine, B., S. Janicot (1996) Near-global sea surface temperature variability associated with West African rainfall anomaly types, *Journal of Climate*, 9, 2935-2940

Fontaine, B., S. Trzaska, S. Janicot (1998) Evolution of the relationship between near global and Atlantic SST modes and the rainy season in West Africa: statistical analyses and sensitivity experiments, *Climate Dynamics*, 14, 353-368

Frías, M. D., S. Herrera, A. S. Cofino, J. M. Gutiérrez (2010) Assessing the skill of precipitation and temperature seasonal forecasts in Spain: windows of opportunity related to ENSO events, *Journal of Climate*, 23, 209-220

Froude, L. S., L. Bengtsson, K. I. Hodges (2013) Atmospheric predictability revisited, *Tellus A*, 65, 19022

Garcia-Morales, M. B., L. Dubus (2007) Forecasting precipitation for hydroelectric power management: how to exploit GCM's seasonal ensemble forecasts, *International Journal of Climatology*, 27, 1691

Ghil, M. (2002) Natural climate variability, *Encyclopedia of global environmental change*, 1, 544-549

Giannini, A., R. Saravanan, P. Chang (2003) Oceanic forcing of Sahel rainfall on interannual to interdecadal timescales, *Science*, 302, 1027-1030

Giannini, A., R. Saravanan, and P. Chang (2005), Dynamics of the boreal summer African monsoon in the NSIPP1 atmospheric model, *Climate Dynamics*, 25, 517-535

Gneiting, T., A. E. Raftery, A. H. Westveld, T. Goldman (2005) Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation, *Monthly Weather Review*, 133, 1098-1118

Gneiting, T., A. E. Raftery (2005) Weather forecasting with ensemble methods, *Science*, 310, 248-249

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, M. A. Cane (2001) Current approaches to seasonal to interannual climate predictions, *International Journal of Climatology*, 21, 1111-1152

Goddard, L., S. Mason (2002) Sensitivity of seasonal climate forecasts to persisted SST anomalies, *Climate Dynamics*, 19, 619-632

Goddard, L., A. Kumar, A. Solomon, D. Smith, G. Boer, P. Gonzalez, V. Kharin, W. Merryfield, C. Deser, S. J. Mason, B. P. Kirtman, R. Msadek, R. Sutton, E. Hawkins, T. Fricker, G. Hegerl, C. A. T. Ferro, D. B. Stephenson, G. A. Meehl, T. Stockdale, R. Burgman, A. M. Greene, Y. Kushnir, M. Newman, J. Carton, I. Fukumori, T. Delworth (2013) A verification framework for interannual-to-decadal predictions experiments, *Climate Dynamics*, 40, 245-272

Graham, R. J., M. Gordon, P. J. McLean, S. Ineson, M. R. Huddleston, M. K. Davey, A. Brookshaw, R. T. H. Barnes (2005) A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model, *Tellus A*, 57, 320-339

Gronos, S. (2005) Vilhelm Bjerknes' vision for scientific weather prediction, *The Nordic Seas: An Integrated Perspective*, 357-366

Grimm, A. M., S. E. Ferraz, J. Gomes (1998) Precipitation anomalies in southern Brazil associated with El Niño and La Niña events, *Journal of Climate*, 11, 2863-2880

Hadley, G. (1735) Concerning the cause of the general trade-winds, *Philosophical Transactions*, 39, 58-62

Hagedorn, R., F. J. Doblas-Reyes, T. N. Palmer (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept, *Tellus A*, 57, 219-233

Halpert, M. S., C. F. Ropelewski (1992) Surface temperature patterns associated with the Southern Oscillation, *Journal of Climate*, 5, 577-593

Hartmann, D. L. (1994) *Global physical climatology*, Academic press

Hartmann, D. L., A. M. G. Klein Tank, M. Rusticucci, L. V. Alexander, S. Brönnimann, Y. Charabi, F. J. Dentener, E. J. Dlugokencky, D. R. Easterling, A. Kaplan, B. J. Soden, P. W. Thorne, M. Wild, P. M. Zhai (2013) Observations: Atmosphere and Surface. In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Hersbach, H. (2000) Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, *Weather and Forecasting*, 15, 559-570

Holton, J. R. (2004) *An Introduction to Dynamic Meteorology*, Elsevier, Burlington

Hoskins, B. (2013) The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science, *Quarterly Journal of the Royal Meteorological Society*, 139, 573-584

Hourdin, F., I. Musat, F. Guichard, P. M. Ruti, F. Favot, M.-A. Filiberti, M. Pham, J.-Y. Grandpeix, J. Polcher, P. Marquet, A. Boone, J.-P Lafore, J-L Redelsperger, A Dell'aquila, T. Losada, A. K. Traore, H. Gallée (2010) AMMA-Model Intercomparison Project, *Bulletin of the American Meteorological* Society, 91, 95-104

Huffman, G. J., D. T. Bolvin (2013) GPCP Version 2.2 Combined Precipitation Data Set Documentation, Internet Publication, 1-46, (Available at: http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.2/doc/V2.2_doc.pdf), Accessed: 16 November 2012

Hurrell, J. W. (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation, *Science*, 269, 676-679

Hurrell, J. W. (2008) Decadal climate prediction: challenges and opportunities, *Journal of Physics: Conference Series*, 125, 012018

Im, E.-S., R. L. Gianotti, E. A. B. Eltahir (2014) Improving the simulation of the West African monsoon using the MIT regional climate model, *Journal of Climate*, 27, 2209-2229

Ingram, K. T., M. C. Roncoli, P. H. Kirshen (2002) Opportunities and constraints for farmers of west Africa to use seasonal precipitation forecasts with Burkina Faso as a case study, *Agricultural Systems*, 74, 331-349

Janicot, S., A. Harzallah, B. Fontaine, V. Moron (1998) West African monsoon dynamics and eastern equatorial Atlantic and Pacific SST anomalies, *Journal of Climate*, 11, 1874-1882

Janicot, S., S. Trzaska, I. Poccard (2001) Summer Sahel-ENSO teleconnection and decadal time scale SST variations, *Climate Dynamics*, 18, 303-320

Johansson, A., A. Barnston, S. Saha, H. van den Dool (1998) On the level and origin of seasonal forecast skill in northern Europe, *Journal of the Atmospheric Sciences*, 55, 103-127

Jolliffe, I. T., D. B. Stephenson (2012) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Second Edition, John Wiley & Sons, Ltd, Chichester

Joly, M., A. Voldoire (2009) Influence of ENSO on the West African Monsoon: Temporal Aspects and Atmospheric Processes, *Journal of Climate*, 22, 3193-3210

Joly, M., A. Voldoire (2010) Role of the Gulf of Guinea in the inter-annual variability of the West African monsoon: what do we learn from CMIP3 coupled simulations?, *International Journal of Climatology*, 30, 1843-1856

Kirtman, B. P., D. Min (2009) Multimodel Ensemble ENSO Prediction with CCSM and CFS, *Monthly Weather Review*, 137, 2908-2930

Kirtman, B., S. B. Power, J. A. Adedoyin, G. J. Boer, R. Bojariu, I. Camilloni, F. J. Doblas-Reyes, A. M. Fiore, M. Kimoto, G. A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G. J. van Oldenborgh, G. Vecchi, H. J. Wang (2013) Near-term Climate Change: Projections and Predictability. In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kirtman, B. P., D. Min, J. M. Infanti, J. L. Kinter III, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, M. Rienecker, M. Suarez, Z. E. Li, J. Marshak, Y.-K. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis, E. F. Wood (2014) The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction, *Bulletin of the American Meteorological Society*, 95, 585-601

Kim, H. M., P. J. Webster, J. A. Curry (2012) Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter, *Climate Dynamics*, 39, 2957-2973

Knutti, R. (2010) The end of model democracy?, *Climatic Change*, 102, 395-404

Koetse, M. J., P. Rietveld (2009) The impact of climate change and weather on transport: An overview of empirical findings, *Transportation Research Part D: Transport and Environment*, 14, 205-221

Korecha, D., A. Sorteberg (2013) Validation of operational seasonal rainfall forecast in Ethiopia, *Water Resources Research*, 49, 7681-7697

Kug, J.-S., J.-Y. Lee, I.-S. Kang (2007) Global sea surface temperature prediction using a multi-model ensemble, *Monthly Weather Review*, 135, 3239-3247

Kug, J.-S., J.-Y. Lee, I.-S. Kang, B. Wang, C.-K. Park (2008) Optimal multi-model ensemble method in seasonal climate prediction, *Asian-Pacific Journal of Atmospheric Sciences*, 44, 259-267

Kumar, K., M. Hoerling, B. Rajagopalan (2005) Advancing dynamical prediction of Indian monsoon rainfall, *Geophysical Research Letters*, 32, L08704

Kumar, A., M. Chen, L. Zhang, W. Wang, Y. Xue, C. Wen, L. Marx, B. Huang (2012) An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2, *Monthly Weather Review*, 140, 3003-3016

Lau, K.-M., S. Yang (2002) *Walker Circulation*. Encyclopedia of Atmospheric Sciences, Academic Press, 2505-2509

Lang, Y., A. Ye, W. Gong, C. Miao, Z. Di, J. Xu, Y. Liu, L. Luo, Q. Duan (2014) Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China, *Journal of Hydrometeorology*, 15, 1546-1559

Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, M. Prather (2007) Historical Overview of Climate Change. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, H. L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Losada, T., B. Rodríguez-Fonseca, S. Janicot, S. Gervois, F. Chauvin, P. Ruti (2010) A multi-model approach to the Atlantic Equatorial mode: Impact on the West African monsoon, *Climate Dynamics*, 35, 29-43

Lorenz, E. N. (1960) The statistical prediction of solutions of dynamic equations. In: *Symposium on Numerical Weather Prediction in Tokyo*

Lorenz, E. N. (1963) Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, 20, 130-141

Lorenz, E. N. (1967) *The nature and theory of the general circulation of the atmosphere*, 218, World Meteorological Organization, Geneva

Lorenz, E. N. (1972) Predictability; Does the Flap of a Butterfly's wings in Brazil Set off a Tornado in Texas?, In: American Association for the advancement of science, 139th meeting

Lorenz, E. N. (1975) Climatic predictability. In: *The Physical Basis of Climate and Climate Modelling*, GARP Publications Series, World Meteorological Organization, 16, 132-136

Lorenz, E. N. (1982) Atmospheric predictability experiments with a large numerical model, *Tellus*, 34, 505-513

Lorenz, E. N. (1983) A history of prevailing ideas about the general circulation of the atmosphere, *Bulletin of the American Meteorological Society*, 64, 730-769

Lynch, P. (1993) Richardson's Forecast factory: the $64 000 question, *Meteorological Magazine*, 122, 69-70

Lynch, P. (2008) The origins of computer weather prediction and climate modeling, *Journal of Computational Physics*, 227, 3431-3444

Manabe, S., K. Bryan (1969) Climate calculations with a combined ocean-atmosphere model, *Journal of the Atmospheric Sciences*, 26, 786-789

Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, L. Sun, P. A. Arkin (1999) The IRI Seasonal Climate Prediction System and the 1997/98 El Niño Event, *Bulletin of the American Meteorological Society*, 80, 1853-1873

Mason, S. J., G. M. Mimmack (2002) Comparison of some statistical methods of probabilistic forecasting of ENSO, *Journal of Climate*, 15, 8-29

Mason, S. J. (2008a) From dynamical model predictions to seasonal climate forecasts, in: *Seasonal Climate: Forecasting and Managing Risk*, Springer Netherlands, 205-234

Mason, S. J. (2008b) Understanding forecast verification statistics, *Meteorological Applications*, 15, 31-40

Mason, S. J., O. Baddour (2008) Statistical modeling. In: Troccoli, A., M. S. J. Harrison, D. L. T. Anderson, S. J. Mason (Ed) *Seasonal Climate: Forecasting and Managing Risk*, Springer Academic Publishers, Dordrecht, 167-206

Mason, S. J., D. B. Stephenson (2008) How can we know whether the forecasts are any good? In: Troccoli, A., M. S. J. Harrison, D. L. T. Anderson, S. J. Mason (Ed) *Seasonal Climate: Forecasting and Managing Risk*, Springer Academic Publishers, Dordrecht, 259-289

McWilliams, J. C. (1996) Modeling the oceanic general circulation, *Annual Review of Fluid Mechanics*, 28, 215-248

Meehl, G. A., T. F. Stocker, W. D. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, A. Noda, S. C. B. Raper, I. G. Watterson, A. J. Weaver, Z.-C. Zhao (2007) Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, H. L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York

Merryfield, W. J., W.-S. Lee, G. J. Boer, V. V. Kharin, J. F. Scinocca, G. M. Flato, R. S. Ajayamohan, J. C. Fyfe, Y. Tang, and S. Polavarapu (2013) The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization, *Monthly Weather Review*, 141, 2910-2945

Mock, D.R. (1981) The Southern Oscillation: Historical Origins. Unpublished term paper, University of Washington, Seattle, WA, Available at: http://www.esrl.noaa.gov/psd/enso/misc/hxsoi.html.

Mohino, E., S. Janicot, J. Bader (2011a) Sahel rainfall and decadal to multi-decadal sea surface temperature variability, *Climate Dynamics*, 37, 419-440

Mohino, E., B. Rodriguez-Fonseca, T. Losada, S. Gervois, S. Janicot, J. Bader, P. Ruti, F. Chauvin (2011b) Changes in the interannual SST-forced signals on West African rainfall: AGCM intercomparison, *Climate Dynamics*, 37, 1707-1725

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, F. Vitart (2011) The new ECMWF seasonal forecast system (System 4), ECMWF Technical Memorandum 656, (Available at http://www.ecmwf.int/publications/library/do/references/list/14), accessed on 20 December 2012

Motha, R. P., S. K. Leduc, L. T. Steyaert, C. M. Sakamoto, N. D. Strommen (1980) Precipitation Patterns in West Africa, *Monthly Weather Review*, 108, 1567-1578

Motter, A. E., D. K. Campbell (2013) Chaos at fifty, *Physics Today*, 66, 27-33

Murphy, A. H., R. L. Winkler (1984) Probability forecasting in meteorology, *Journal of the American Statistical Association*, 79, 489-500

Murphy, A. H. (1991) Forecast verification: Its Complexity and Dimensionality, *Monthly Weather Review*, 119, 1590-1601

Murphy, J. M., D. M. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, D. A. Stainforth (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768-772

Nicholls, N. (2001) Commentary and analysis: The Insignificance of Significance Testing, *Bulletin of the American Meteorological Society*, 82, 981-986

Nicholson, S. E. (1993) An overview of African rainfall fluctuations of the last decade, *Journal of Climate*, 6, 1463-1466

Palmer, T. N., D. L. T. Anderson (1994) The prospects for seasonal forecasting - A review paper, *Quarterly Journal of the Royal Meteorological Society*, 120, 755-793

Palmer, T. N. (2000) Predicting uncertainty in forecasts of weather and climate, *Reports on Progress in Physics*, 63, 71-116

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson (2004) Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER), *Bulletin of the American Meteorological Society*, 85, 853-872

Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, M. Leutbecher (2005) Representing model uncertainty in weather and climate prediction, *Annual Review of Earth and Planetary Sciences*, 33, 163-193

Panagiotopoulos, F., M. Shahgedanova, D. B. Stephenson (2002) A review of Northern Hemisphere winter-time teleconnection patterns, *Journal de Physique IV*, 12, 27-47

Phillips, N. A. (1956) The general circulation of the atmosphere: A numerical experiment, *Quarterly Journal of the Royal Meteorological Society*, 82, 123-164

Philippon, N., F. J. Doblas-Reyes, P. M. Ruti (2010) Skill, reproducibility and potential predictability of the West African monsoon in coupled GCMs, *Climate Dynamics*, 35, 53-74

Pielke Sr, R. A. (1998) Climate prediction as an initial value problem, *Bulletin of the American Meteorological Society*, 79, 2743-2746

Rasmusson, E. M., J. M. Wallace (1983) Meteorological aspects of the El Nino/Southern Oscillation, *Science*, 222, 1195-1202

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research*, 108, 4407-4444

Richardson, L. (1922) *Weather Prediction by Numerical Process*, Cambridge, The University press

Robertson, A. W., U. Lall, S. E. Zebiak, L. Goddard (2004) Improved combination of multiple atmospheric GCM ensembles for seasonal prediction, *Monthly Weather Review*, 132, 2732-2744

Rodó, X., E. Baert, F. A. Comin (1997) Variations in seasonal rainfall in Southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation, *Climate Dynamics*, 13, 275-284

Rodrigues, L. R. L., F. J. Doblas-Reyes, C. A. S. Coelho (2014a) Multi-model calibration and combination of tropical seasonal sea surface temperature forecasts, *Climate Dynamics*, 42, 597-616

Rodrigues, L. R. L., J. García-Serrano, F. J. Doblas-Reyes (2014b) Seasonal forecast quality of the West African monsoon rainfall regimes by multiple forecast systems, *Journal of Geophysical Research: Atmospheres*, 119, 7908-7930

Rodríguez-Fonseca, B., S. Janicot, E. Mohino, T. Losada, J. Bader, C. Caminade, F. Chauvin, B. Fontaine, J. García-Serrano, S. Gervois, M. Joly, I. Polo, P. Ruti, P. Roucou, A. Voldoire (2011) Interannual and decadal SST-forced responses of the West African Monsoon, *Atmospheric Science Letters*, 12, 67-74

Roehrig, R., D. Bouniol, F. Guichard, F. Hourdin, J.-L. Redelsperger (2013) The Present and future of the West African monsoon: A processoriented assessment of CMIP5 simulations along the AMMA transect, *Journal of Climate*, 26, 6471-6505

Ropelewski, C. F., M. Halpert (1987) Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation, *Monthly Weather Review*, 115, 1606-1626

Roulston, M. S., L. A. Smith (2002) Weather and seasonal forecasting, In: Dischel, R. (ed.) *Climate risk and the weather market*, Risk Books, London, 115-126

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, H. L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng, P. Xie (2006) The NCEP Climate Forecast System, *Journal of Climate*, 19, 3483-3517

Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, E. Becker (2014) The NCEP Climate Forecast System Version 2, *Journal of Climate*, 27, 2185-2208

Saji, H. N., B. N. Goswami, P. N. Vinayachandran, T. Yamagata (1999) A dipole mode in the tropical Indian Ocean, *Nature*, 401, 360-363

Scaife, A. A., A. Arribas, E. Blockley, A. Brookshaw, R. T. Clark, N. Dunstone, R. Eade, D. Fereday, C. K. Folland, M. Gordon, L. Hermanson, J. R. Knight, D. J. Lea, C. MacLachlan, A. Maidens, M. Martin, A. K. Peterson, D. Smith, M. Vellinga, E. Wallace, J. Waters, A. Williams (2014) Skillful long-range prediction of European and North American winters, *Geophysical Research Letters*, 41, 2514-2519

Schneider, T. (2006) The general circulation of the atmosphere, *Annual Review of Earth and Planetary Sciences*, 34, 655-688

Schneider, U., A. Becker, A. Meyer-Christoffer, M. Ziese, and B. Rudolf (2011) Global Precipitation Analysis Products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 1-13, available at: http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU4/KU42/en/Reports__Publications/GPCC__intro__products__2011,templateId=raw,property=publicationFile.pdf/GPCC_intro_products_2011.pdf), accessed: 16 November 2012

Shukla, J. (1998) Predictability in the midst of chaos: A scientific basis for climate forecasting, *Science*, 282, 728-731

Slingo, J., T. N. Palmer (2011) Uncertainty in weather and climate prediction, *Philosophical Transactions of the Royal Society A*, 369, 4751-4767

Sooraj, K. P., H. Annamalai, A. Kumar, H. Wang (2012) A comprehensive assessment of CFS seasonal forecasts over the tropics, *Weather Forecasting*, 27, 3-27

Smith, T. M., R. W. Reynolds, T. C. Peterson, J. Lawrimore (2008) Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006), *Journal of Climate*, 21, 2283-2296

Stephenson, D. B., C. A. S. Coelho, F. J. Doblas-Reyes, M. Balmaseda (2005) Forecast Assimilation: A unified framework for the combination of multi-model weather and climate predictions, *Tellus A*, 57, 253-264

Stephenson, D. B., C. A. S. Coelho, I. T. Jolliffe (2008) Two extra components in the Brier Score decomposition, *Weather and Forecasting*, 23, 752-757

Stephenson, D. B., M. Collins, J. C. Rougier, R. E. Chandler (2012) Statistical problems in the probabilistic prediction of climate change, *Environmetrics*, 23, 364-372

Stockdale, T. N., D. L. T. Anderson, M. A. Balmaseda, F. J. Doblas-Reyes, L. Ferranti, K. Mogensen, T. N. Palmer, F. Molteni, F. Vitart (2011) ECMWF seasonal forecast system 3 and its prediction of sea surface temperature, *Climate Dynamics*, 37, 455-471

Sultan, B., S. Janicot (2000) Abrupt shift of the ITCZ over West Africa and intra-seasonal variability, *Geophysical Research Letter*, 27, 3353-3356

Sultan, B., S. Janicot, A. Diedhiou (2003) The West African monsoon dynamics. Part I: documentation of intraseasonal variability, *Journal of Climate*, 16, 3389-3406

Sultan, B., S. Janicot (2003) The West African monsoon dynamics. Part II: The "preonset" and "onset" of the summer monsoon, *Journal of Climate*, 16, 3407-3427

Sewell, W. R. D., R. W. Kates, L. Phillips (1968) Human Response to Weather and Climate: Geographical Contributions, *Geographical Review*, 58, 262-280

Sylla, M. B., I. Diallo, J. S. Pal (2013) West African Monsoon in State-of-the-Science Regional Climate Models, in Climate Variability - Regional and Thematic Patterns, available at: http://www.intechopen.com/books/climate-variability-regional-and-thematic-patterns/west-african-monsoon-in-state-of-the-science-regional-climate-models

Tippet, M. K., A. Giannini (2006) Potentially predictable components of African summer rainfall in an SST-forced GCM simulation, *Journal of Climate*, 19, 3133-3144

Tippett, M. K., A. G. Barnston (2008) Skill of Multimodel ENSO Probability Forecasts, *Monthly Weather Review*, 136, 3933-3946

Trenberth, K. E. (1997) The definition of El Niño, *Bulletin of the American Meteorological Society*, 78, 2771-2777

Trenberth, K. E., K. Miller, L. Mearns, S. Rhodes (2000) *Effects of changing climate on weather and human activities*, University science books, Sausalito, California

Trenberth, K. E., D. J. Shea (2006) Atlantic hurricanes and natural variability in 2005, Geophysical Research Letter, 33, L12704

Trenberth, K. E., P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden, P. Zhai (2007) Observations: Surface and Atmospheric Climate Change. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York

Tompkins, A. M., L. Feudale (2010) Seasonal ensemble predictions of West African monsoon precipitation in the ECMWF System 3 with a focus on the AMMA special observing period in 2006, *Weather and Forecasting*, 25, 768-788

Vellinga, M., A. Arribas, R. Graham (2013) Seasonal forecasts for regional onset of the West African monsoon, *Climate Dynamics*, 40, 3047-3070

Vernieres, G., M. M. Rienecker, R. Kovach, C. L. Keppenne (2012) The GEOS-iODAS: Description and evaluation, Technical Report Series on Global Modeling and Data Assimilation, 30, 1-61

Walker, G. T., E. W. Bliss (1932) World weather V, *Memoirs of the Royal Meteorological Society*, 4, 53-84

Wallace, J. M., D. S. Gutzler (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Monthly Weather Review*, 109, 784-812

Wallace, J. M., P. V. Hobbs (2006) *Atmospheric science: an introductory survey*, 92, Academic press

Wang, C. (2005) ENSO, Atlantic climate variability, and the Walker and Hadley circulations, In: *The Hadley circulation: Present, past and future*, Springer Netherlands, 173-202

Wang, B., J.-Y. Li, I.-S. Kang, J. Shukla, C.-K. Park, A. Kumar, J. Schemm, S. Cocke, J.-S. Kug, J.-J. Luo, X. Fu, W.-T. Yun, O. Alves, E. Jin, J. Kinter, B. Kirtman, T. Krishnamurti, N. Lau, W. Lau, P. Liu, P. Pegion, T. Rosati, S. Schubert, W. Stern, M. Suarez, T. Yamagate (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14 model ensemble retrospective seasonal prediction (1980-2004), *Climate Dynamics*, 33, 93-117

Wanner, H., S. Brönnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, E. Xoplaki (2001) North Atlantic Oscillation-concepts and studies, *Surveys in geophysics*, 22, 321-381

Weisheimer, A., T. N. Palmer, F. J. Doblas-Reyes (2011) Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles, *Geophysical Research Letters*, 38, L16703

Wilks, D. (2006) *Statistical methods in the atmospheric sciences*, Second edition, International Geophysics Series, 59, Elsevier, Oxford

Yuan, X., E. F. Wood, L. Luo, M. Pan (2011) A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophysical Research Letter*, 38, L1340

Yoshikatsu, Y., M. Koki, T. Hiroshi (2008) Global Warming Projections Using the Community Climate System Model, CCSM3, *NEC Technical Journal*, 3, 73-76

Zanchettin, D., S. W. Franks, P. Traverso, M. Tomasino (2008) On ENSO impacts on European wintertime rainfalls and their modulation by the NAO and the Pacific multi-decadal variability described through the PDO index, *International Journal of Climatology*, 28, 995-1006

Zhang, S., M. J. Harrison, A. Rosati, A. Wittenberg (2007) System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies, *Monthly Weather Review*, 135, 3541-3564

Zebiak, S. E. (1993) Air-sea interaction in the equatorial Atlantic region, *Journal of Climate*, 6, 1567-1586

Zuo, Z., S. Yang, Z.-Z. Hu, R. Zhang, W. Wang, B. Huang, F. Wang (2013) Predictable patterns and predictive skills of monsoon precipitation in Northern Hemisphere summer in NCEP CFSv2 reforecasts, *Climate Dynamics*, 40, 3071-3088

Appendix A.

Figure A.1 shows the correlation with the observations of the hindcasts produced by the statistical model for the Niño3.4 index in retroactive and cross-validation modes for the period between 1982 and 2010. Similar patterns are found, including the decrease in correlation (blue squares) for 4-6 month lead predictions produced early in the year, particularly during the northern hemisphere spring. This feature is known as the spring barrier (Balmaseda et al., 1995). The correlation is also very similar. The main differences are found in the boreal summer (June-July-August) for leads longer than four months. In these cases, the statistical model in retroactive mode performs better (i.e. has higher correlation) than the statistical model trained in cross-validation mode, which justified the use of the former approach in the results shown in this paper. It has been shown in previous studies that the predictive skill estimation methods using simple or multiple regression in cross-validation mode could introduce negative bias in negative correlations (Barnston and Van den Dool 1993).

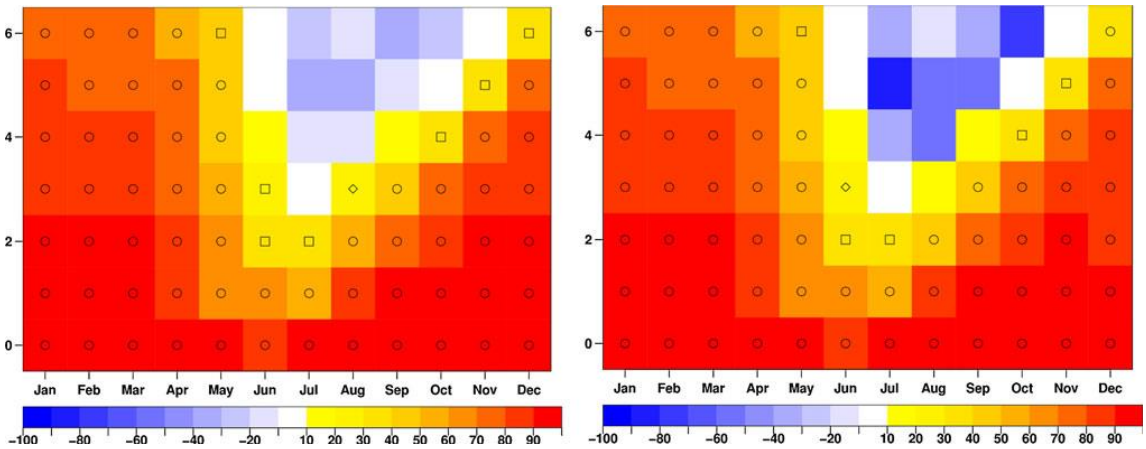

Figure A.1: Correlation between the predicted and observed Niño3.4 index as a function of target month (horizontal axis) and lead time (vertical axis) for the statistical model trained in forecast mode (left column) and in cross-validation mode (right column). Predictions have been formulated over the period 1982–2010. HadISST data are used to estimate the coefficients in the statistical model and for the forecast quality assessment. The symbols are for the p values (see text for details). Circles are for p values smaller than or equal 0.01, squares for p values between 0.05 and 0.01, and diamonds for p-values between 0.10 and 0.05

Appendix B.

Figure B.1 shows the correlation coefficient between indices of the Guinean and Sahelian regimes and three SST indices for all months of the year for the period 1951-2011. SST indices representing the main SST variability over several ocean regions are obtained via spatial averaging. These indices represent the main patterns of climate variability and are widely used as predictive tools in statistical models (Doblas-Reyes et al., 2013). The Equatorial Pacific, North and Equatorial Atlantic ocean basins are known to play an important role on the WAM rainfall variability (Folland et al., 1986; Fontaine and Janicot, 1996; Fontaine et al., 1998; Joly and Voldoire, 2009, 2010; Mohino et al., 2011a, 2011b; Rodríguez-Fonseca et al., 2011). Therefore, the Niño3.4 (SST anomalies averaged over 170ºW-120ºW; 5ºS-5ºN), the Atlantic Multidecadal Oscillation (AMO; SST anomalies averaged over 80ºW-0ºW; 0º-60ºN minus global SST anomalies over 60ºS-60ºN) and the Atlantic 3 (Atl3; SST anomalies averaged over 20ºW-0ºW; 3ºS-3ºN) indices are used. The SST over other regions such as the Mediterranean basin that might also play a role on the WAM rainfall variability (Fontaine et al., 2010), is not taken into account for the sake of simplicity. The correlation between the rainfall regimes and the SST indices is computed using the period 1951-2011. The Niño3.4 SST index is not well correlated either with the Guinean or the Sahelian regime (the maximum absolute correlation values are 0.31 and 0.25, respectively). This might be either because the Niño3.4-WAM rainfall relationship is not stationary (Mohino et al., 2011b; Rodríguez-Fonseca et al., 2011) or because not all ENSO events can be linked to WAM rainfall anomalies (Joly and Voldoire, 2009). The time series associated with the AMO (black line) and the Atl3 (blue line) are almost of opposite sign when comparing the Guinean (left panel) and the Sahelian (right panel) regimes. As shown previously, there is positive correlation between the Atl3 and the Guinean regime (Joly and Voldoire, 2010) and the AMO and the Sahelian regime (Mohino et al., 2011a; Rodríguez-Fonseca et al., 2011). Therefore, we used the Atl3 and the AMO as predictors for the Guinean and Sahelian regimes, respectively. The PCs associated with the WAM rainfall regimes are computed for the target months between June and October (see Section 2 for detailed information) and, as a consequence, only the months prior to June of the target year may be considered as predictors when trying to mimic an operational forecasting approach. Figure B.1 shows that the best predictor for the Guinean regime is the Atl3 of May of the target year while the best predictor for the Sahelian regime is the AMO of December of the year prior to the target year.

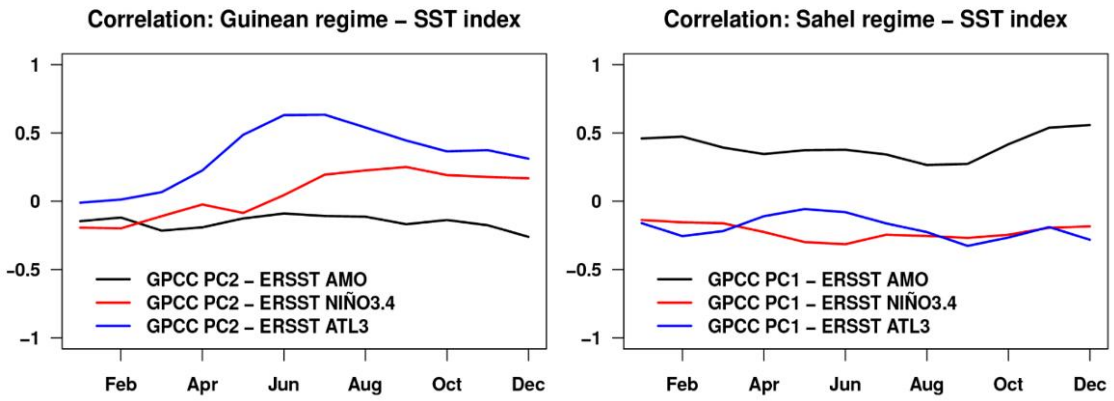

Figure B.1: Correlation coefficient between the Guinean and Sahelian regimes (estimated from the GPCC seasonal evolution diagram described above) and three ERSSTv3b SST indices: AMO, Niño3.4 and Atl3. The correlation is computed for each month of the year and for the period 1951-2011.

Appendix C.

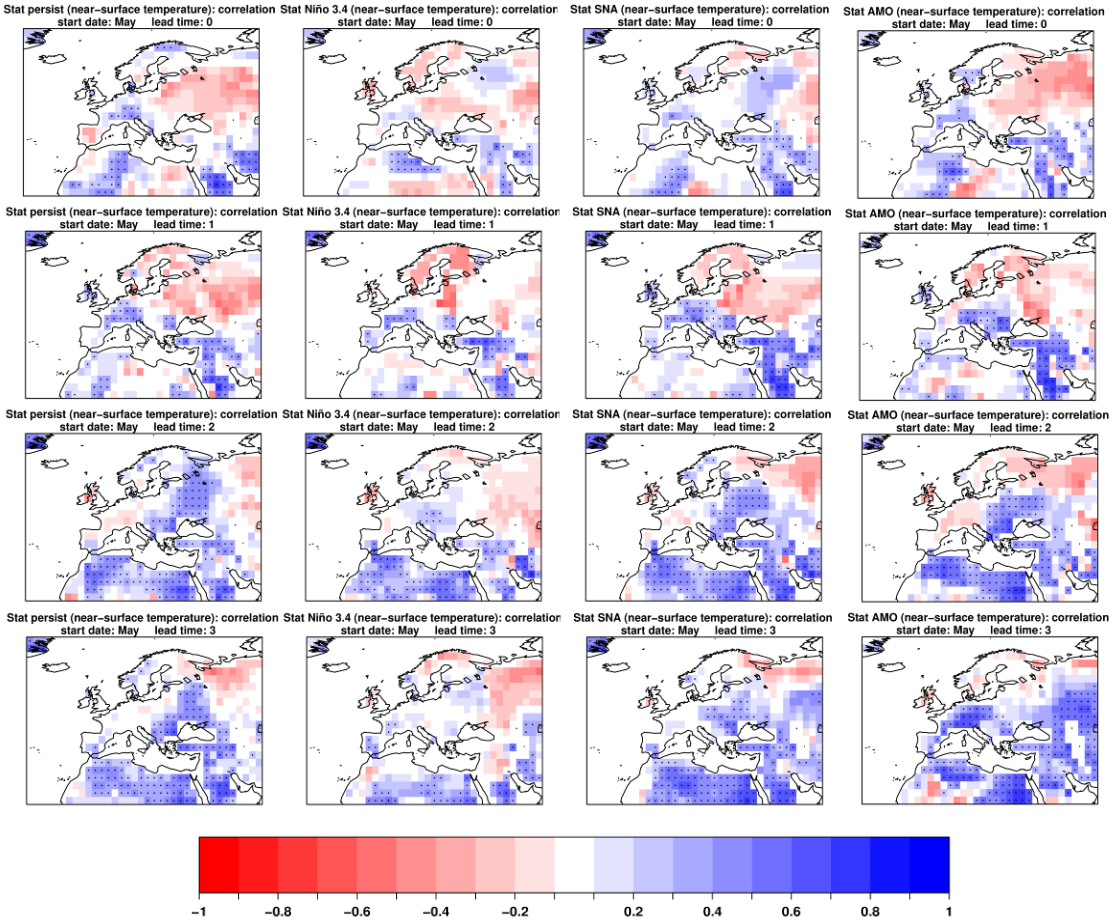

Figure C.1: Correlation between predicted and observed near-surface temperature in May (first row), June (second row), July (third row) and August (fourth row). Anomaly values of April is used to predict the summer months of May (lead time zero) through August (lead time three months). The correlation was computed for the hindcast period 1982-2010. The statistical model was estimated using four different predictors: the predictand variable itself at the same grid point (first column), SST Niño3.4 (second column), SNA (third column) and AMO (fourth column) indices. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

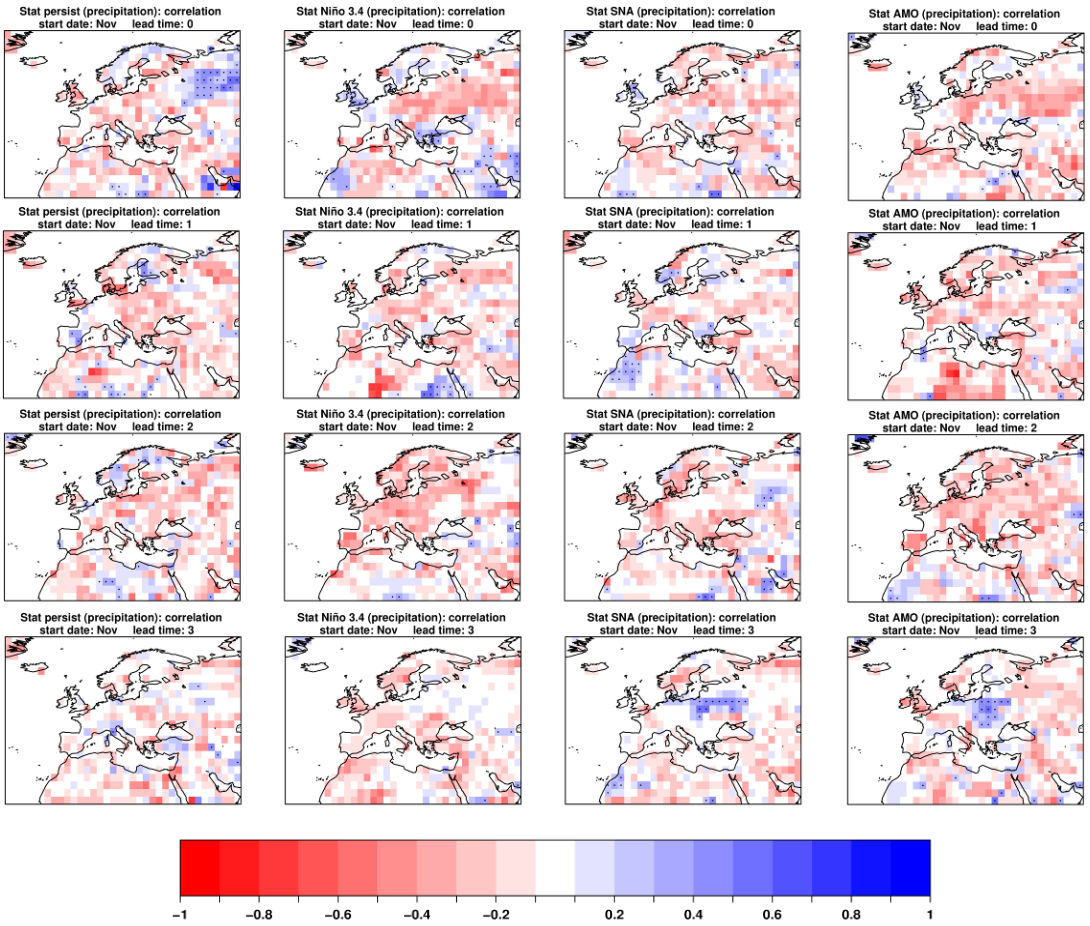

Figure C.2: Correlation between predicted and observed precipitation in November (first row), December (second row), January (third row) and February (fourth row). Anomaly values of November is used to predict the winter months of November (lead time zero) through February (lead time three months). The correlation was computed for the hindcast period 1982-2010. The statistical model was estimated using four different predictors: the predictand variable itself at the same grid point (first column), SST Niño3.4 (second column), SNA (third column) and AMO (fourth column) indices. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

Appendix D.

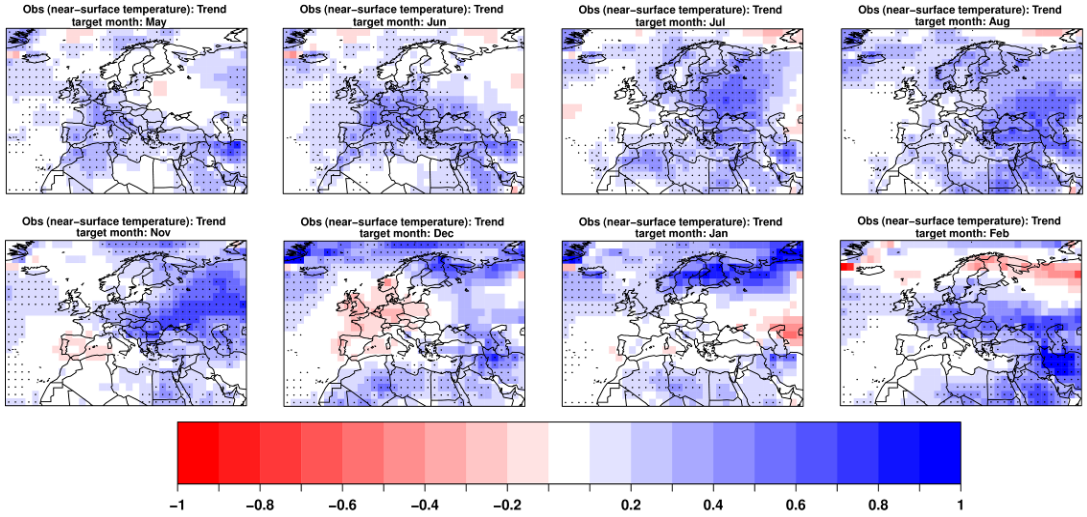

Figure D.1: Observed near-surface temperature linear trend in the summer (May, June, July and August; first row) and winter (November, December, January and February; second row) months computed for the period 1982-2010. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

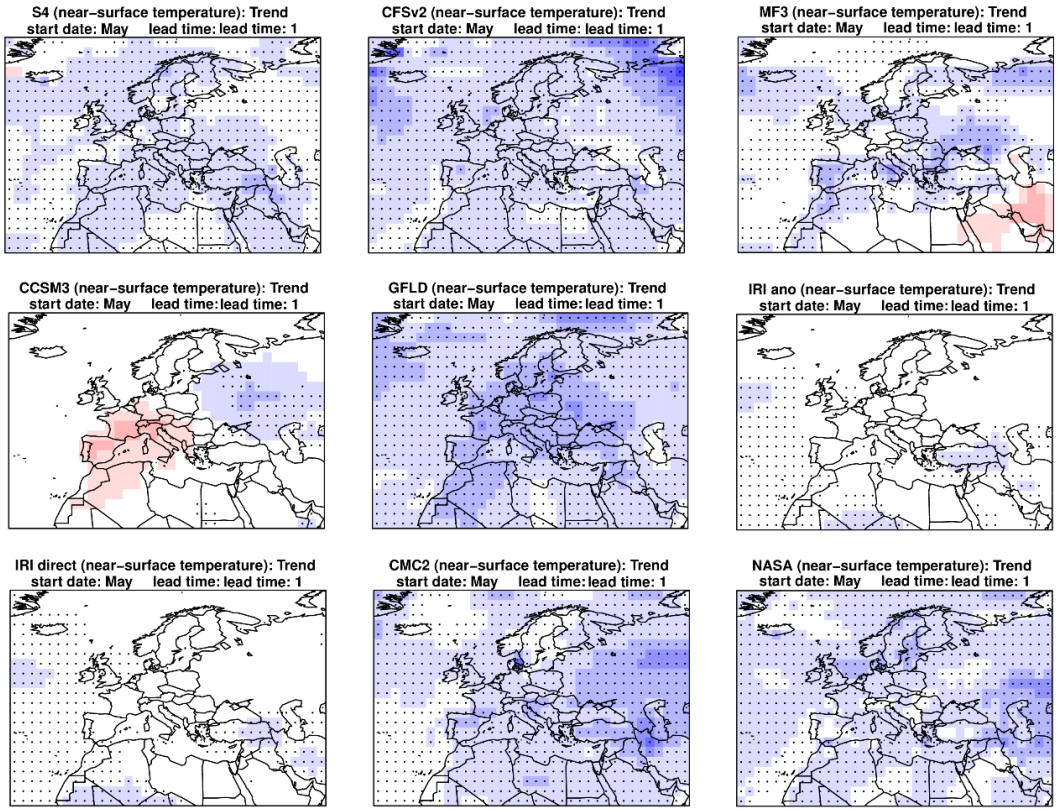

Figure D.2: Predicted near-surface temperature linear trend in June computed for the period 1982-2010. Predictions were initialized in May (lead time one month). The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

Appendix E.

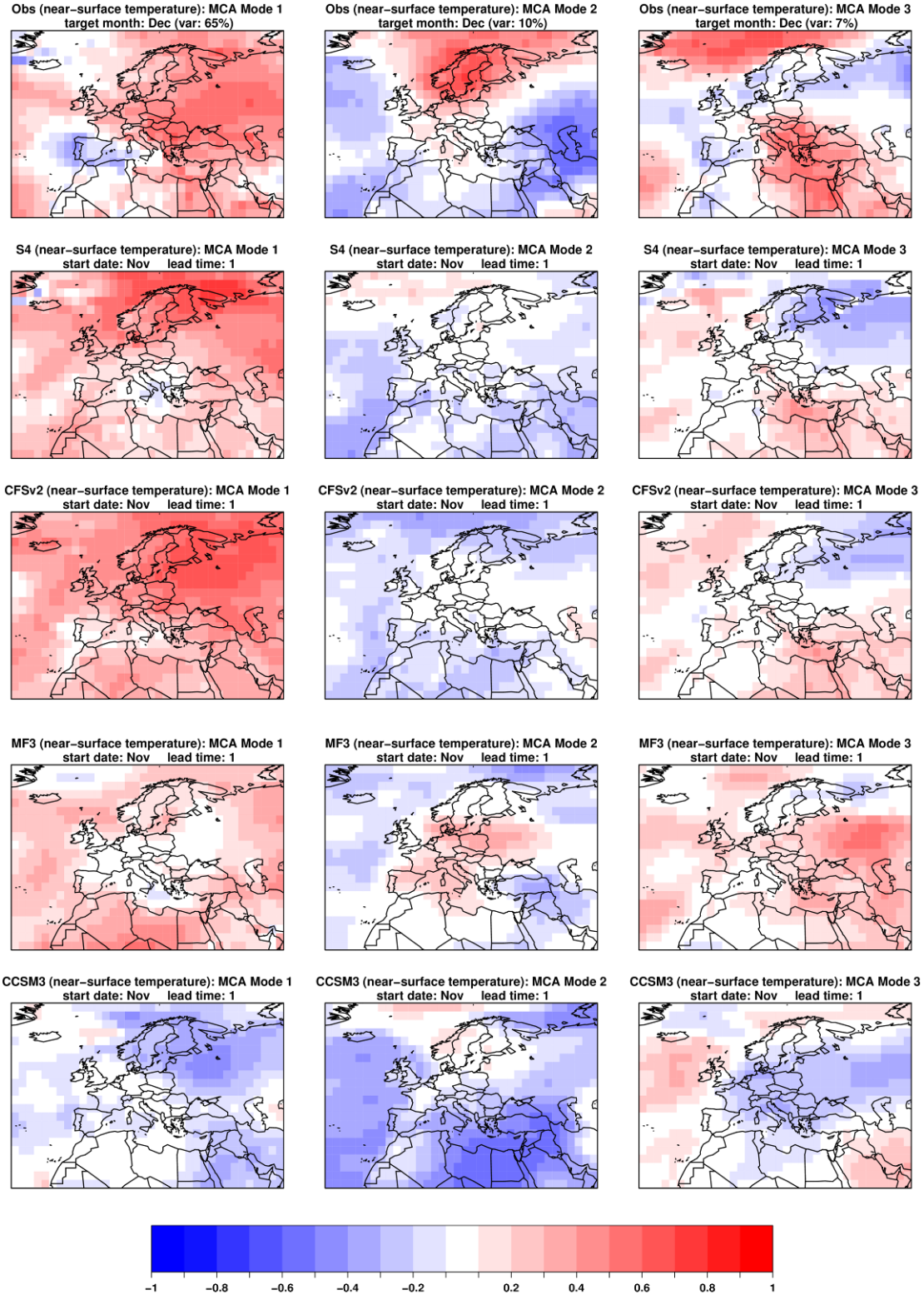

Figure E.1: Observed and predicted heterogeneous correlation maps of near-surface temperature in December. Predictions are for November start date (lead time 1). The expansion coefficients of the left field (i.e. the observation) are correlated with the original data of the right field (i.e. the forecast systems) and vice versa. Results are shown for the first three leading MCA modes, computed for the hindcast period 1982-2010.

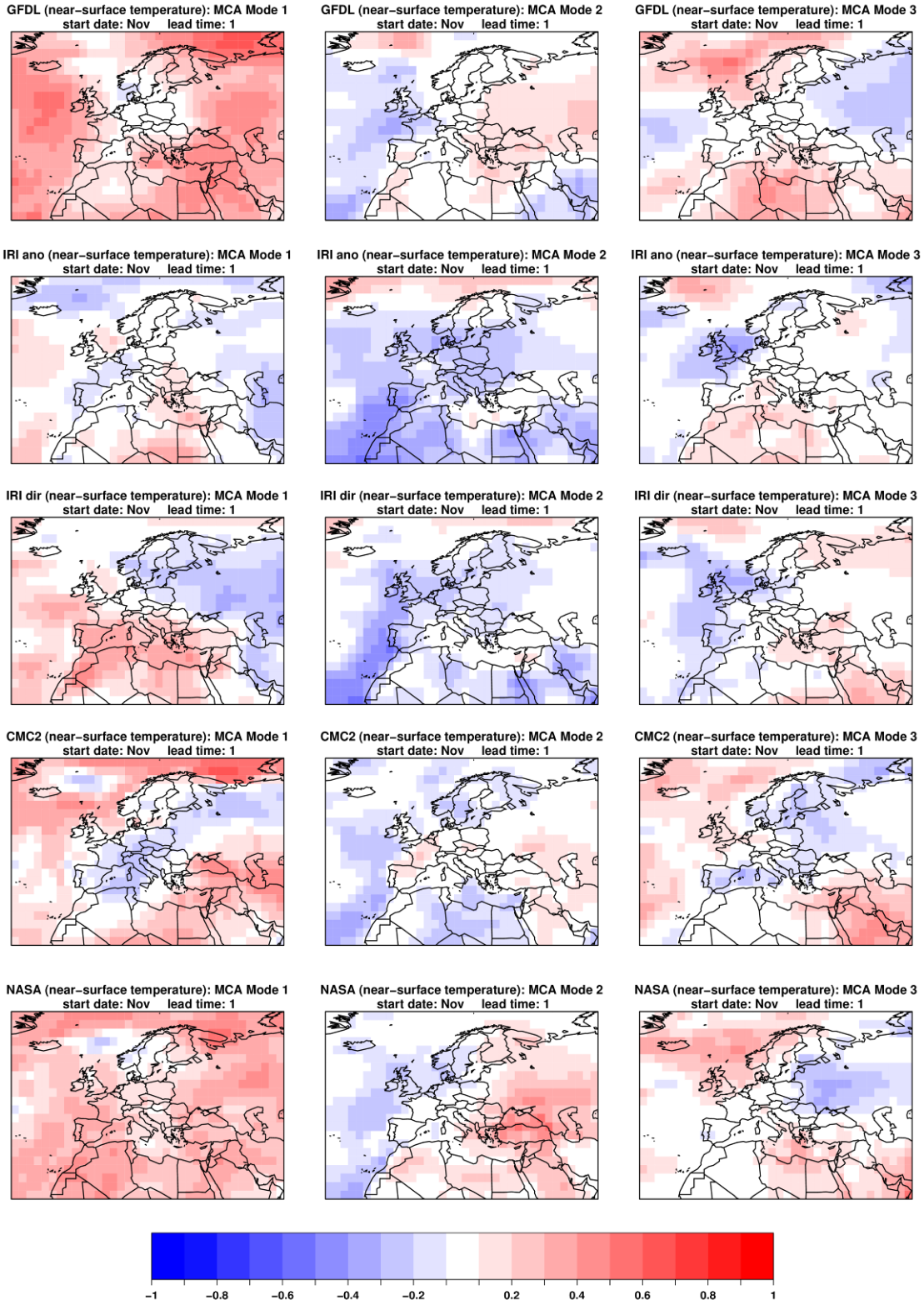

Figure E1: Continue.

Appendix F.

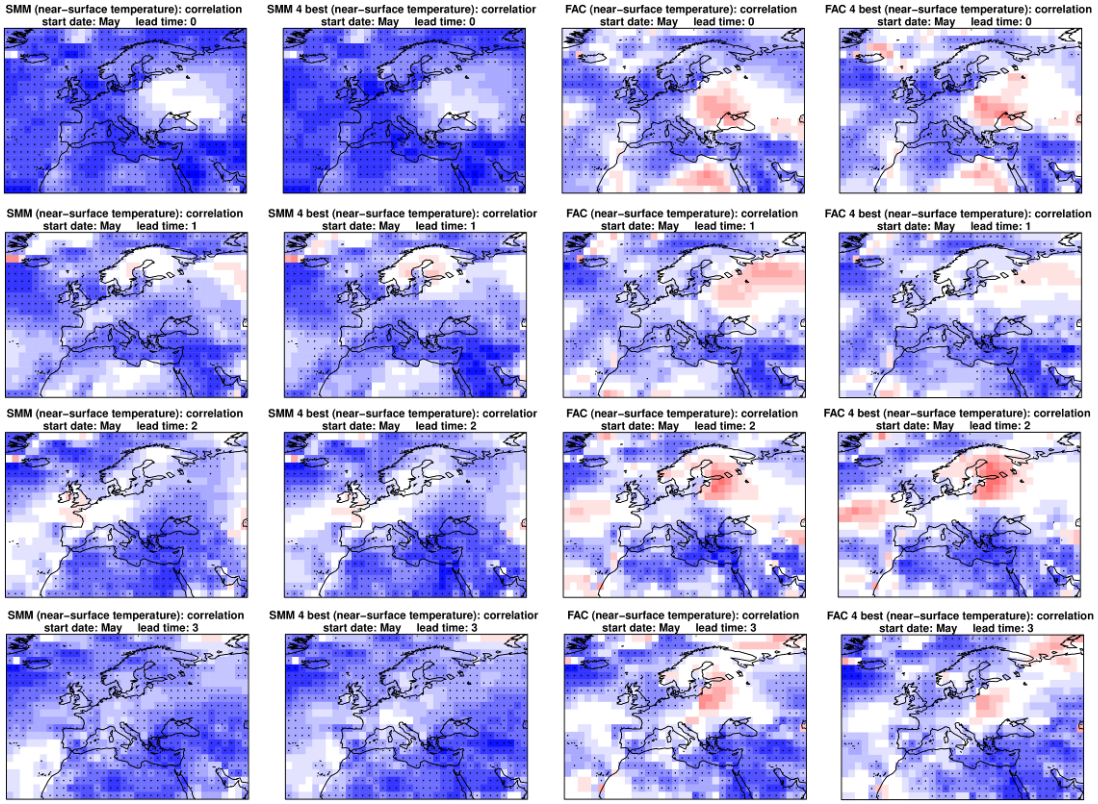

Figure F.1: Correlation coefficient between predicted and observed near-surface temperature in May (first row), June (second row), July (third row) and August (fourth row). Predictions are for May start dates (lead times zero through three months). The correlation was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

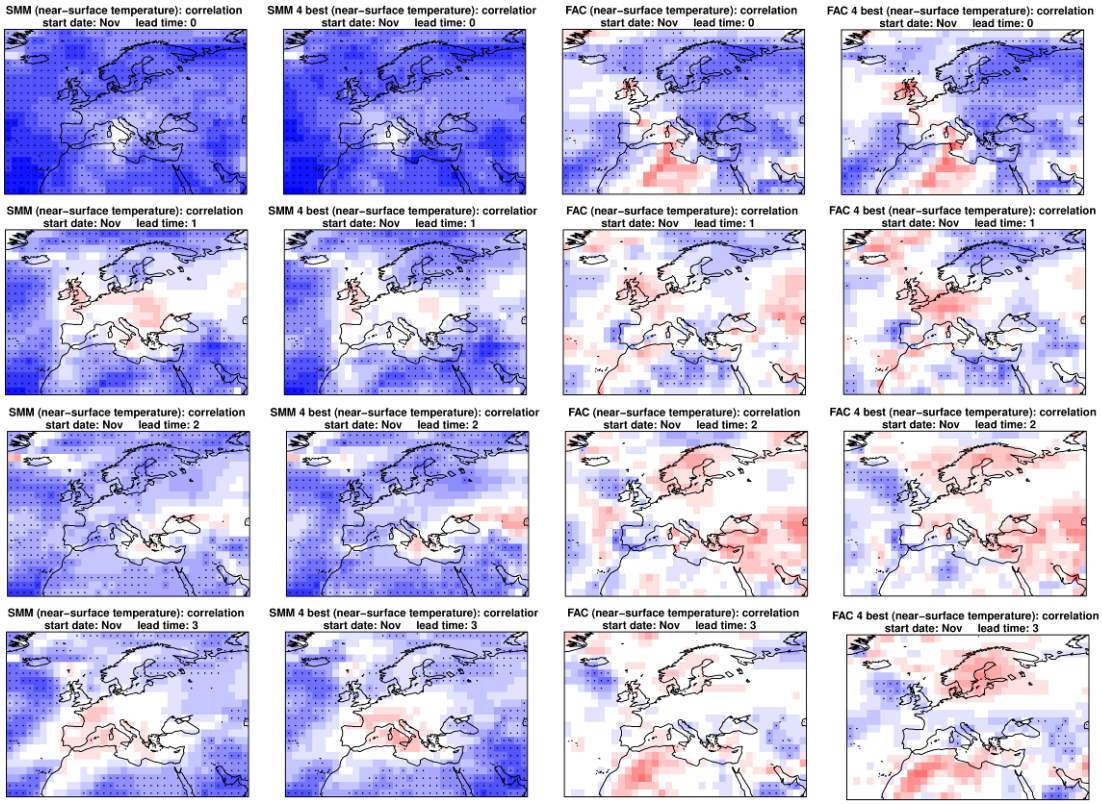

Figure F.2: Correlation coefficient between predicted and observed near-surface temperature in November (first row), December (second row), January (third row) and February (fourth row). Predictions are for November start dates (lead times zero through three months). The correlation was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

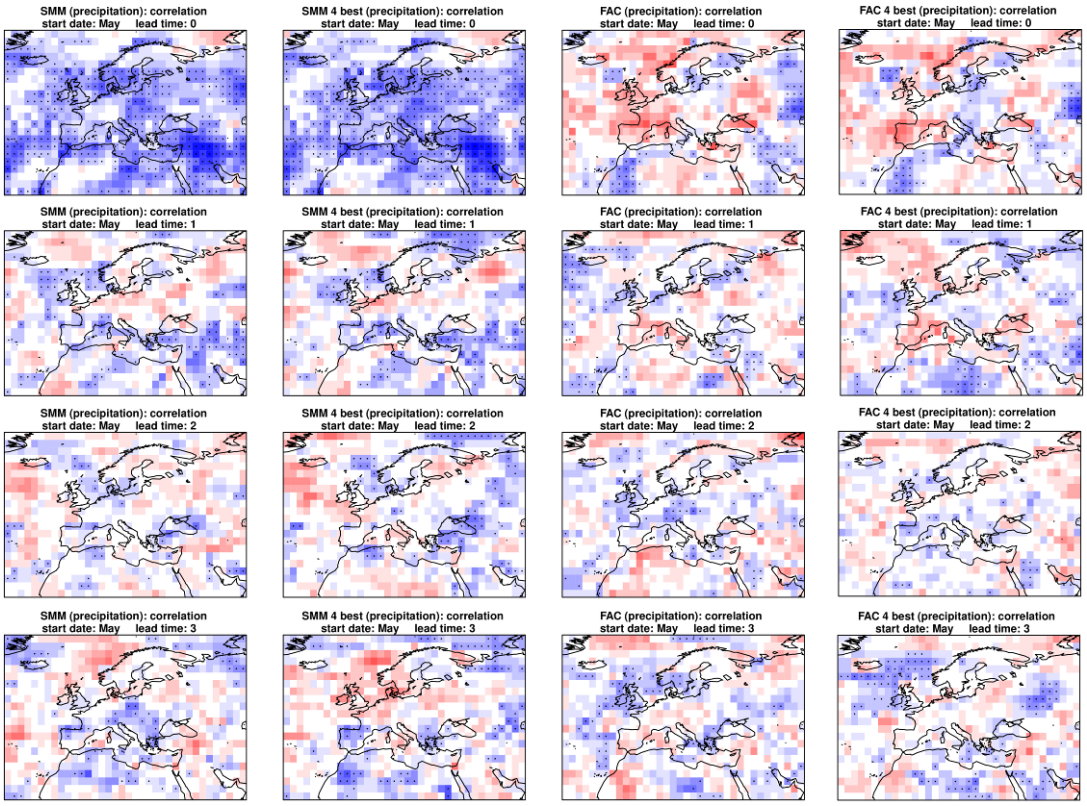

Figure F.3: Correlation coefficient between predicted and observed precipitation in May (first row), June (second row), July (third row) and August (fourth row). Predictions are for May start dates (lead times zero through three months). The correlation was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

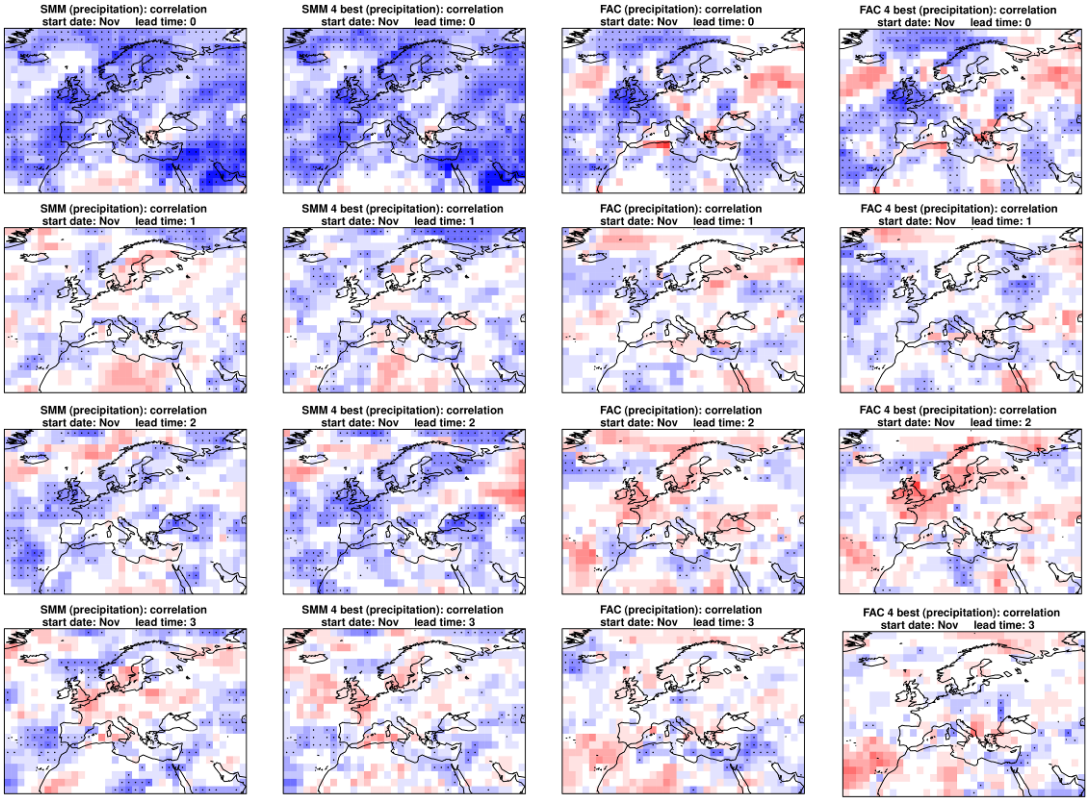

Figure F.4: Correlation coefficient between predicted and observed precipitation in November (first row), December (second row), January (third row) and February (fourth row). Predictions are for November start dates (lead times zero through three months). The correlation was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

Appendix G.

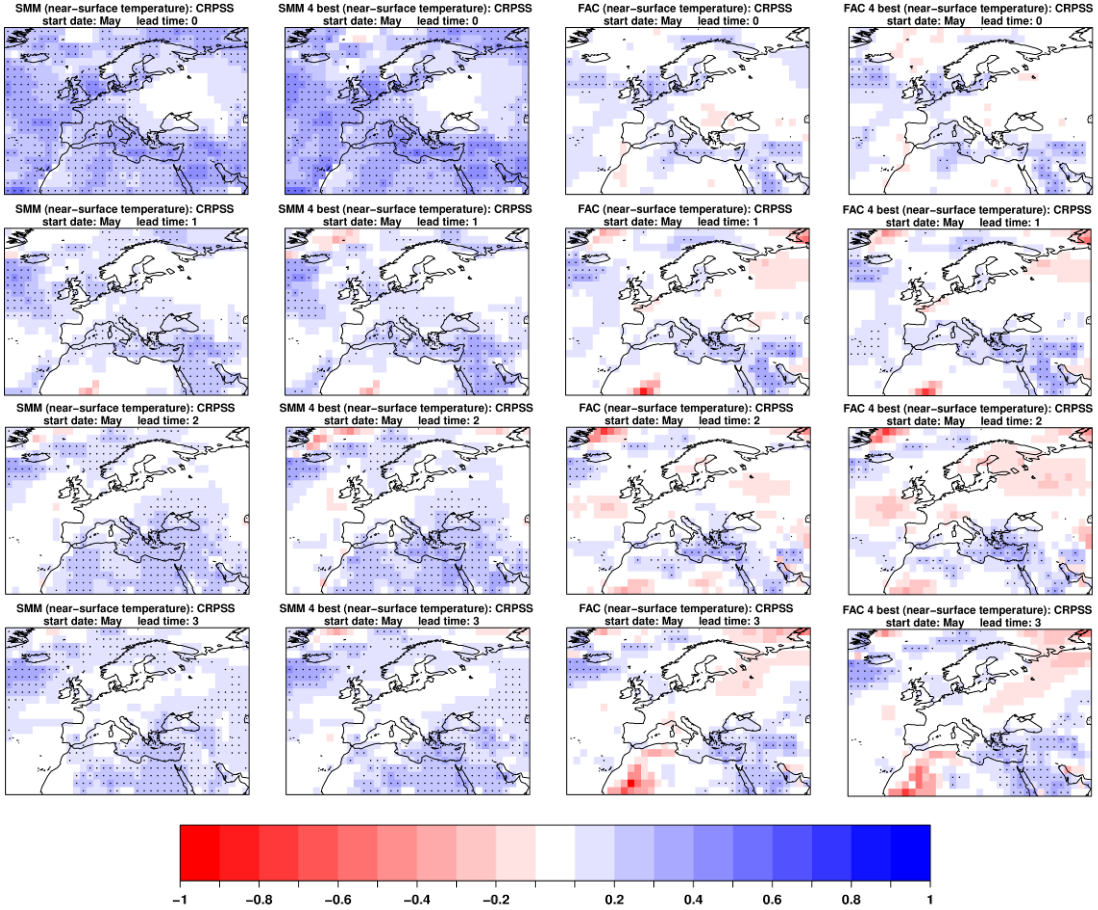

Figure G.1: CRPSS for near-surface temperature predictions in May (first row), June (second row), July (third row) and August (fourth row). Predictions are for May start dates (lead times zero through three months). The CRPSS was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.

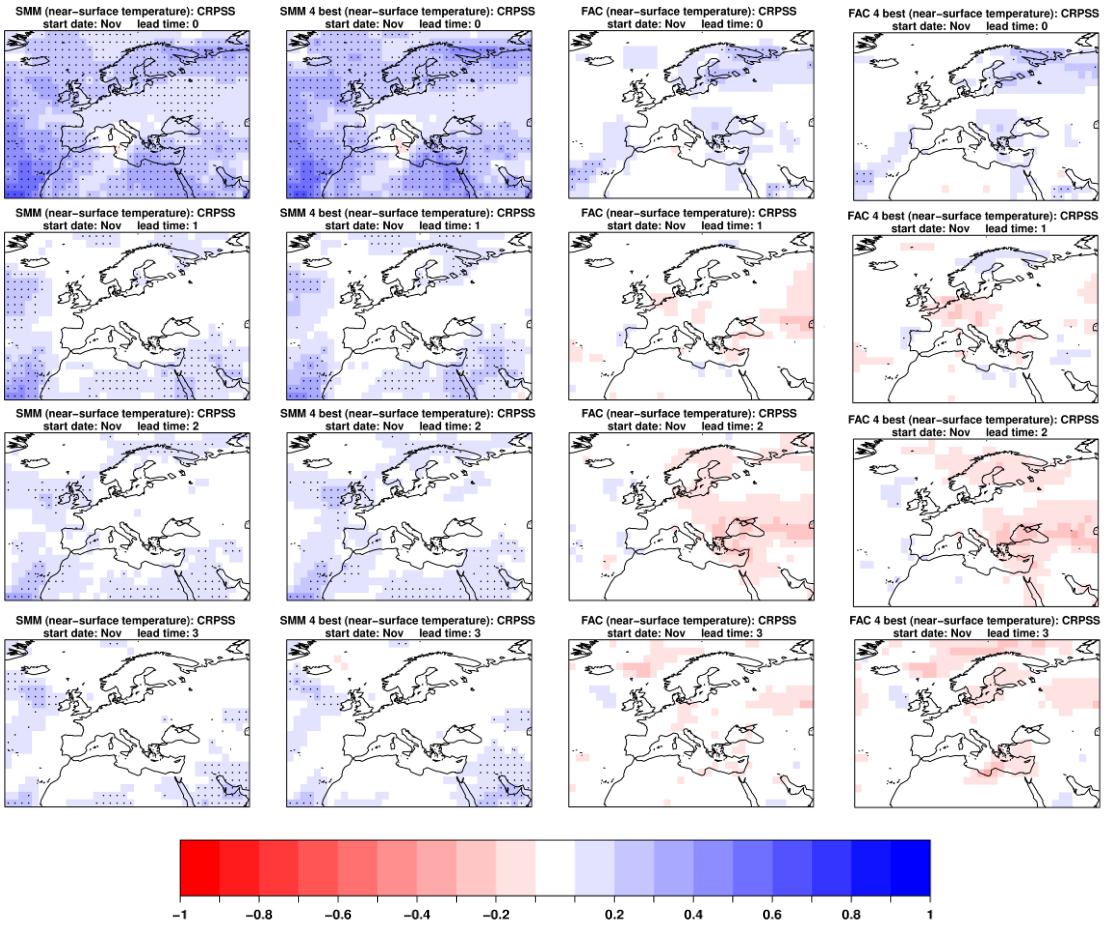

Figure G.2: CRPSS for near-surface temperature predictions in November (first row), December (second row), January (third row) and February (fourth row). Predictions are for November start dates (lead times zero through three months). The CRPSS was computed for the hindcast period 1982-2010. The four combinations are SMM, SMM 4 best, FAC and FAC 4 best and the forecast systems used in the 4 best combinations are S4, CFSv2, GFDL and CMC2. The dots are placed where there is statistical significance at the 95% level computed using non-parametric bootstrap. See text for details.